RADAR ENGINEERING

The quality of the materials used in the manufacture of this book is governed by continued postwar shortages.

RADAR ENGINEERING

by Donald G. Fink

Editor, Electronics; formerly Staff Member Radiation Laboratory, M.I.T., and Expert Consultant Office of the Secretary of War; author of "Principles of Television Engineering," "Engineering Electronics"; editor "Television Standards and Practice"

FIRST EDITION

NEW YORK AND LONDON

McGRAW-HILL BOOK COMPANY, INC.

1947

Preface

DURING the war years, technical workers in the fields of radio engineering and physics were engaged in one of the greatest coordinated efforts in the history of applied science: the development and production of radar systems for military use. Techniques that might have required 20 years of normal development were perfected in less than 5 years, thanks to long hours of intensive effort by virtually all qualified personnel, backed by unlimited technical and fiscal resources.

The body of new information thus acquired could not be published while the war continued, nor could it be exchanged between workers in different parts of the field. Thus it has come about that many workers have specialized knowledge of certain of the new techniques, but few have a broad view embracing all the significant advances. This compartmentation of knowledge, however necessary it may have been during the war, constitutes a serious hindrance to the future advance of the radio art. Recognizing this fact, the Army, Navy, and Office of Scientific Research and Development have declassified nearly all of the hitherto restricted information on radar methods and equipment. Several thousand OSRD reports may be consulted at the Library of Congress and other depots; but these reports are not indexed or correlated, nor are they readily available in any event. Consequently the need for information remains largely unsatisfied.

This book is intended to provide a general compilation of radar information in a single volume. The book is divided into two parts that may be considered, respectively, theoretical and practical. The first part deals with fundamental concepts essential to an understanding of radar technology. These include pulse generation and transmission, waveguides and transmission lines, resonant cavities, the formation of radio-frequency beams, propagation of ultrahigh-frequency and superhigh-frequency signals, and the reflection of radio energy. Nearly all this material was available, in one form or another, before the war. But it was not, and to a large extent still is not, a usual part of the

radio engineer's background. Since a basic foundation in these concepts is essential to an understanding of radar practice, the author has endeavored to present each subject comprehensively, from first principles to the working equations on which practical designs are based. Knowledge of the more familiar aspects of radio science is assumed, including mathematics and electricity of college grade, conventional electron tubes and circuits, audio-frequency and radio-frequency systems. Previous experience with television, pulse techniques, ultra-high frequencies, or microwaves is not assumed.

The second part of the book is devoted to components, circuits, and structures used in radar equipment. Descriptions of typical military radar systems are included. These have been selected, from among those free of security restrictions at the time of writing, to illustrate technical matters, rather than military applications. Thus all the equipments described are Army radars intended for use of ground troops (because these were the first declassified), but the radio frequencies employed include the wide range of 200, 600, 3,000, and 10,000 megacycles.

The question of documentation has proved a difficult one, largely because the art has developed without benefit of the editorial clergy. When the manuscript was completed there were fewer than 25 references, in the periodical literature, dealing specifically with radar methods and equipment. Since that time, during the editing and typesetting, many additional papers have made their appearance. The footnotes and bibliographies have accordingly been brought up to date as of the time of going to press.

The author acknowledges gratefully the assistance of the following companies in supplying photographs to illustrate the text: Western Electric Company, Figs. 268, 269*a*, 277, 284; Raytheon Manufacturing Company, Fig. 269*b*; Westinghouse Electric Corporation, Figs. 350, 352, 428; and the Sperry Gyroscope Company, Figs, 437*b*, 438, 439.

DONALD G. FINK.

TENAFLY, N.J.,
January, 1947.

Contents

PAGE

Preface vii

PART I—Radar Fundamentals

CHAPTER I

Introduction to Radar Concepts 3

Origins of Radar—Historical Survey—Methods of Radar Detection—Pulse System—Continuous-wave System—Technical Specifications of a Pulse Radar System—Pulse Specifications—Radio-frequency Specifications—Scanner Specifications—Circular Scanning—Helical Scanning—Conical Scanning—Spiral Scanning—Lobe Switching—The Indicator Specifications—Type A Indicator—Type B Indicator—Type C Indicator—Plan-position Indicator—Modifications of Type A Indicator—Modifications of Type B Indicator—Modifications of Type C Indicator—Summary of Indicator Types—The Radar Target—Target Types—Distinction between Types of Target—Background Returns—Terrain Echoes—The Free-space Radar Equation—Influence of Operating Wavelength—Influence of Receiver Characteristics—Numerical Examples Based on the Free-space Radar Equation—Summary

CHAPTER II

Principles of Pulse Generation and Transmission 63

Pulse Methods in Radar—Transient Response of Circuit Elements—Transient Response of Single-circuit Elements—Resistance—Transient Response of Pairs of Circuit Elements—Resistance and Inductance in Series—Transient Response of Pairs of Elements in Shunt—Circuits of Three or More Elements—Transient Response of Vacuum Tubes—Grid-voltage Cutoff—Grid-current Cutoff—Transient Effect of Tube Capacitances—Transient Operation of Diodes—Steady-state Analysis of Pulse Circuits—Determination of Pulse Spectra—Spectra of Repetitive Rectangular Pulse—The Fourier Integral—Computation of System Functions—Single-Terminal Pair Networks—System Functions of the Video Coupling Connection—System Functions of Two-terminal-pair Network—System Functions of the Ideal Low-pass Filter—Steady-state Computation of Output Waveforms—Pulse Transmission by the Video Amplifier Coupling Connection—Pulse Transmission by the Ideal Low-pass Filter—Relative Effects of Amplitude and Phase Distortion—Transmission of Pulse-modulated Radio-frequency Carriers—Noise in Pulse Systems—Bandwidth for Maximum Signal-to-noise Ratio—Bandwidth Criterion for High Resolution—The Noise Concept in Radar

CHAPTER III

PAGE

Radio-frequency Fundamentals—Transmission Lines, Waveguides, and Resonant Cavities 134

Radio-frequency Techniques in Radar—Radio Frequencies Employed in Radar—Transmission Lines—Mechanism of Transmission—Characteristic and Terminating Impedances—Loss-free Lines—Impedance and Admittance Circle Diagrams—Circle Diagram of the Loss-free Line—The Admittance Circle Diagram and Shunt Reactive Elements—The Smith Circle Diagram—Power Transmission and Measurement in Transmission Lines—Parameters of Transmission Lines in Terms of Dimensions and Materials—Waveguides, General Considerations—Mechanism of Propagation in Waveguides—Group Velocity, Phase Velocity, and Wavelength in Guide—Solution of Transverse Electric Modes in Rectangular Guides—Modes in Circular Waveguide—Attenuation in Waveguides—Choice of Mode and Dimensions of Waveguides—Excitation and Termination of Waveguides—Reflecting Elements for Impedance Matching in Waveguides—Oscillating Circuits and Resonant Cavities—Energy Storage and Dissipation in Resonant Cavities—Modes in Cavity Resonators—Rectangular Form—Excitation of Cavity Resonators

CHAPTER IV

Radio-frequency Fundamentals (Continued)—Radiators and Reflectors, Propagation, and Targets 219

Forms of Radiating Structures—Properties of Infinitesimal Oscillating Dipole—Properties of the Half-wave Linear Radiator—Waveguide Radiators—Directive Properties of Dipole Arrays—Directivity of Uniform-current Dipole Arrays—Single-sided Radiation from Arrays—Reduction of Side Lobes by Current Distribution—Lobe Shifting by Phase Relationships—Physical Forms of Dipole Arrays—Arrays of Discrete Half-wave Elements—Paraboloidal Reflectors—Paraboloidal Sections—Absorptive Properties of Directive Radiators—Formation of Beams by Electromagnetic Lenses—Propagation of Radar Signals—Propagation over Flat Reflecting Earth—Propagation over Spherical Reflecting Earth—Anomalous Propagation of Radar Signals—Refractive Bending of Wave Paths—Computation and Measurement of Modified Index Curves—Relation of *M* Curves to Weather Phenomena—Absorption of Signals by the Atmosphere—The Radar Target—Target Contrast—Effect of Target Surroundings—Targets in Motion—Doppler Effect

Part II—RADAR CIRCUITS AND COMPONENTS

CHAPTER V

Introduction to Radar Design 293

Radar Characteristics and Specifications—Technical Specifications—The Radiated Beam—Scanning Specifications—Choice of Radio Fre-

PAGE

quency—Pulse Specifications—Receiver Specifications—Indicator Specifications—Technical Specifications of Representative Military Radars—Selection of Circuits and Components

CHAPTER VI

Basic Pulse Circuits 311

Functions of Pulse Circuits—Basic Timing Sources—Direct (Unsynchronized) Generation of Pulses and Extended Waveforms—The Multivibrator—The Blocking Oscillator—Synchronized Production of Pulses from Extended Waveforms—Deformation of Sinewaves by Limiters—The Differentiator Circuit—Synchronized Production of Pulses—Synchronized Multivibrators and Blocking Oscillators—Pulse-forming Transmission Lines—Saturable Core Reactors—Synchronized Production of Extended Waveforms—Synchronized Production of Sawtooth Waves—Modification of Waveform Shape—Pulse-rate Multiplication and Division—Amplification of Pulse Waveforms—Linear Video Amplification—Compensation of Video Amplifiers—High-frequency Compensation—Low-frequency Performance—Current and Power Relations in Video Amplifiers—Cathode-coupled Video Amplifiers—The Bootstrap Circuit—The Cathode-follower Circuit—Pulse Transformers—Time-delay Circuits—Phase Shift of Sinusoidal Sources—Artificial Transmission Lines—Delay Multivibrators—Pulse Circuit Auxiliaries—Maintenance of Direct-current Reference—Regulated Power Supply

CHAPTER VII

Basic Radio-frequency Circuits and Structures 359

Coaxial Lines, Waveguides, and Associated Fittings—Support of Coaxial Inner Conductors—Rotating Joints in Coaxial Lines—Waveguide Joints—Radio-frequency Generators (Transmitting Oscillators)—External Line Triode Oscillators—Ultrahigh-frequency Oscillator Circuits—The Cavity Magnetron—Internal Action of the Cavity Magnetron—Operating Characteristics of the Cavity Magentron—The Rieke Diagram—Duplex Switching Systems (Transmit-receive Boxes)—Transmit-receive Switches for Microwave Systems—Radiating Structures—Lobe Switching—Radiators for Conical Scanning—High-speed Scanners—Nondirectional Radiators—Radio-frequency Amplifiers—Superheterodyne Components—Local Oscillators—The Reflex Klystron—Condition for Maximum Output in a Reflex Klystron—Practical Local Oscillator Circuits—Heterodyne Frequency Converters (Mixers)—Silicon Crystals—Microwave Mixers

CHAPTER VIII

Synchronization Equipment (Timers) 431

Functions of Timing Equipment—Self-synchronous Timers—Rotary Spark Gaps—Auxiliary Circuits—Master Oscillator Timers—Sinewave Oscillator Operating at Pulse Frequency—Master Oscillator Timers—Crystal-controlled Timing for Precision Range Measurements—Block Diagram of SCR-584 Timer—Simplified Schematic of SCR-584 Timer—Other Timing Methods

PAGE

CHAPTER IX

Transmitters and Radiators 458

Functions and Requirements of the Modulator—Types of Modulator Circuit Elements—Basic Modulator Circuits—Driver Circuits—Typical Transmitters and Radiators—Type SCR-268—The AN/-TPS-3 Transmitter and Radiator (500 Megacycle Rotary-gap Equipment)—The SCR-584 Transmitter and Radiator (3,000 Megacycle Gunfire-control Equipment)—The AN/MPG-1 Transmitter and Radiator System

CHAPTER X

Receivers 496

Receiver Functions and Components—Noise in Receivers—Choice of Amplification Method—Radio-frequency Components—Intermediate-frequency Amplifier Design—Video Detection and Amplification—Tuning Auxiliaries—Typical P and L Band Radar Receivers—Types SCR-268 and AN/TPS-3—Typical Microwave Receivers—SCR-584 and AN/MPG-1

CHAPTER XI

Indicators and Scanners 534

Components and Functions of Scanners and Indicators—The Cathode-ray Indicator Tube—Phosphors—Deflection Systems—Auxiliary Circuits for Cathode-ray Tubes—Deflection Patterns for Various Types of Indication—Linear Sweeps for Range Measurement—Plan-position-indicator Sweep Systems—Direction Sweep Potentiometers—Data-transmission Systems—Servomechanisms and Power Drives—Typical Indicator and Scanner Systems—SCR-268—Scanner and Indicator of the AN/TPS-3 Radar—Scanner and Indicator of the SCR-584—Scanner and Indicator of the AN/MPG-1 Radar

CHAPTER XII

Radio-frequency Measurements and Test Equipment . . . 590

Ultrahigh-frequency and Superhigh-frequency Measurements—Bench Oscillators—Power-measuring Devices—Attenuators and Fixed Loads—Frequency Standards and Wavemeters—Standing-wave Detectors—Tuners, Couplers, and Power Dividers—Measurement of Q—Spectrum Analyzer

Index 627

PART I

Radar Fundamentals

CHAPTER I

INTRODUCTION TO RADAR CONCEPTS

The term "radar" was coined by the U. S. Navy in 1941 to describe equipment for "*ra*dio *d*etection *a*nd *r*anging," the word being formed from the letters italicized. The word is applied, by common usage, to radio equipment used to detect the presence of objects by means of reflected radio waves.

There are two categories of radar systems: the pulse radars and the c-w (continuous-wave) radars. In the pulse equipments, sharp bursts of radio energy, somewhat like the bursts of acoustic energy from the barrel of a machine gun, are sent out. When these bursts, or "pulses," encounter a reflecting object, they are reflected as discrete echoes, which are detected by the radar receiver during the interval between the transmitted pulses.

The second type of equipment, the c-w radar, sends out a continuous flow of radio energy. The reflected wave is distinguished from the outgoing signal by a slight change in radio frequency. This change in frequency can be imposed on the transmitted wave by frequency modulation, or it may arise from relative motion between the radar and the reflecting object.

The pulse method has proved to be more adaptable to military needs, particularly in its ability to measure distances and to engage several targets simultaneously. As a result, pulse radar equipment has been developed, produced, and utilized almost to the exclusion of c-w radar. The c-w method has many interesting properties, particularly its ability to distinguish moving targets against a stationary reflecting background, which have accounted for its limited use. Moreover, the fact that the c-w system is more conservative of ether space than the pulse system may enhance its prospects in the future, when such space may be limited.

1. Origins of Radar.—Although the development of radar as a weapon of war goes back little more than a decade, the general principles have been known and used for many years. A particularly remarkable application of the principle, which occurs

in nature, is the sound-ranging apparatus used by bats to avoid obstacles in nocturnal flight.[1] While in flight, bats emit from the larynx sharp pulses of supersonic acoustic energy, the carrier frequency being 50 kc, and the pulses occurring at a rate up to 50 per second. The ears, their directivity accentuated by the short (0.7-cm) wavelength employed, detect and time the echoes of the bursts and thus permit the bat to avoid obstacles. An elementary form of the same ability is acquired by blind persons, who experience increased difficulty in avoiding obstacles when the sense of hearing is cut off.

Technically, radar has marked similarity to other devices such as the sonic echo depth finder and detector, which found wide use in World War I. The presence of radio reflections has been appreciated and understood from the very beginnings of radio, and the reflection method was actively employed as a scientific tool in ionosphere studies as early as 1924.

The development of radar, as distinguished from the preceding ionosphere studies, arose from military necessity. The bomber plane had outstripped the methods of detecting it, that is, the searchlight and the sound locator. The searchlight had limited range at best; it was useless in the presence of clouds, and it revealed its own presence. The sound locator lost out because the speed of flight is an appreciable part of the speed of sound. High-flying aircraft move an appreciable distance before the sound reaches the earth. The position indicated by the sound locator is, therefore, always behind the plane, and the actual position must be predicted on the assumption of straight flight. Such assumptions are vitiated by evasive action on the part of the pilot.

Radar solved these problems at a stroke. The range, even in the earliest equipment, is greater than that of the searchlight. The radar indicates where the plane is at the instant it is detected, since the speed of radio waves is several million times as great as that of the plane. The radio waves are not seriously obstructed by clouds, fog, or precipitation. Their presence is unknown to

[1] The study that revealed this fact was carried out in Cruft Laboratory, Harvard University, by Robert Galambos and D. R. Griffin, using supersonic equipment developed by G. W. Pierce. See GALAMBOS, ROBERT, Flight in the Dark; A Study of Bats, *Sci. Monthly*, **56**, 155–162 (February, 1943); and GRIFFIN, D. R., and R. GALAMBOS, *Jour. Exp. Zool.*, **86**, 481–506 (1941).

the enemy, unless he carries receiving equipment to search the radio spectrum.

The concept of military radar arose in the period 1934–1936. Since that time radar has been developed, by all belligerents, on a scale never before equalled in any weapon of comparable use. In 10 years radar has become an essential part of aerial and naval warfare. Now that hostilities are concluded, not only is the employment of radar in peacetime pursuits of great potential value, but the techniques derived from radar bid fair to revolutionize many of the arts of electrical communication.

2. Historical Survey.—The invention of radar cannot be ascribed to a particular person or organization. Rather it is necessary, and appropriate, to trace its development in terms of concepts advanced by a succession of workers. In chronological order, the essential discoveries are as follows: (1) that radio waves are reflected by objects possessing dielectric constant, permeability, or conductivity different from that of their surroundings: (2) that radio waves may be directed along narrow beams; (3) that distance may be measured by timing the travel of a radio wave; (4) that reflected waves may be detected by wave interference patterns; (5) the pulse method of measuring distances; and (6) the concept of combining these potentialities in a device specifically to give knowledge of enemy activity. Not until the last step had been taken was radar invented in a narrowly legal sense. But the dependence of this last step on all the preceding ones is so obvious that credit for the invention must be assigned to all participants.

The earliest credit is easy to assign: it belongs to the first radio physicist, Heinrich Hertz. Hertz, testing Maxwell's hypothesis that light and heat were different forms of ether vibration, described in 1887 experiments[1] in which he proved that radio waves were reflected like light rays, and could be formed into beams by metallic mirrors similar in shape to the mirrors used to form beams of light. To Hertz, if not to Maxwell himself, goes the credit for the discovery of two of radar's essentials.

These experiments were repeated many times in the following years, and there was general appreciation among informed radio

[1] HERTZ, HEINRICH, *Ann. Physik*, **36**, 769 (1889). Hertz demonstrated, with parabolic mirrors, the effects of reflection, polarization, and refraction, using a wavelength of 66 cm.

workers that reflections were present in radio propagation. The relationship between time of travel and distance was abundantly demonstrated in the very act of measuring the speed of light waves. Strangely enough the speed of radio wave propagation has never been measured with the same precision as that of light. It has been necessary to calculate the small difference between the two values in terms of ground and atmospheric constants. The basic identification of time with distance in radio wave propagation goes back to Hertz's proof that radio and light followed the same natural laws.

The fourth concept, that radio reflections could be detected by wave interference, was also clear to many workers. An early record of direct observation of this effect is contained in a confidential report written in 1922 by Dr. A. Hoyt Taylor and Mr. L. C. Young, of the U.S. Naval Research Laboratory, at Anacostia, D.C. These men, whose position in the development of radar is outstanding, noted that signals received on 60 megacycles were subject to occasional rapid variation in signal strength, and finally identified the cause as the passage of boats on the near-by Anacostia River. At this time, Dr. Taylor ascribed the effect to an interference pattern and suggested that surface vessels might in fact be discovered by this means. This observation antedates other radar work by a number of years.

In the meantime the direct measurement of distance by the employment of c-w and pulsed signals had progressed in the field of ionosphere research. The ionosphere had been postulated in 1902 by Kennelly[1] and Heaviside[2] and its existence was first proved by Sir Edward Appleton and Barnett[3] in England in 1924, using a frequency-shift c-w method to determine the approximate height. Anomalies in these results gave rise to the suspicion that the layer was in fact stratified, with several sublayers at different heights. The c-w method, then as now, proved incapable of giving easily identified distinctions between these layers. It was largely as a means of studying the fine structure of the ionosphere that the pulse method, another basic of radar, was developed. In 1925 the first organized program of ionosphere

[1] KENNELLY, A. E., *Elec. World*, **39**, 473 (Mar. 15, 1902).

[2] HEAVISIDE, O., Encyclopaedia Britannica, 10th ed., vol. 9., p. 215, 1902.

[3] APPLETON and BARNETT, *Proc. Roy. Soc.* (*London*), **109**, 621–641 (Dec. 1, 1925).

research, using pulses, began at the Carnegie Institution under Drs. Breit and Tuve.[1] Dr. Taylor and Gebhardt and Young participated in setting up the apparatus. An excellent historical summary of this and other early work has been prepared by Tuska.[2]

In 1930, Taylor resumed his researches at the U.S. Naval Research Laboratory and began a definite program leading to the final concept, a device to detect the enemy and give knowledge of his movements. Aircraft were detected by c-w signals on 60 megacycles, using the wave interference pattern as an indication. In 1930, Taylor suggested the use of such wave interference patterns as a means of detection for military purposes. In 1932, the Secretary of the Navy communicated all information available at that time to the Secretary of War, in the belief that radar would prove essential to antiaircraft activities in the Army.

Up to this time, the detection was based on the c-w principle, which, as stated earlier, did not prove fruitful for military purposes. In 1934, a c-w source of radio waves at frequencies about 3,000 megacycles (10-cm wavelength) and about ½ watt power was tested by the Army Signal Corps Laboratories, and interference patterns from waves reflected from harbor traffic were detected at distances less than a mile. Infrared and heat detectors were actively investigated at the same time. They detected shipping at ranges up to 2 miles.

In July, 1934, Major Blair, director of the Army Signal Corps Laboratories, reported that consideration was being given to "the scheme of projecting an interrupted sequence of trains of oscillations against the target and attempting to detect the echoes during the interstices between the projections." This was one of the first proposals to use pulses in a radio detecting device. No apparatus was built at the time, however. Early in 1935, Sir Robert Watson-Watt of the National Physical Laboratory in England made the same proposal independently, and in May of that year, pulse equipment had been constructed by Watson-Watt and his coworkers and had tracked aircraft.

By December, 1936, Army Signal Corps research on the project had turned wholly to the pulse system, and work began on the

[1] BREIT and TUVE, *Phys. Rev.*, **28** (2d series), 554–575 (September, 1926).

[2] TUSKA, C. D., *Jour. Franklin Inst.*, **237**, 1–20, 83–102 (January and February, 1944).

first U.S. Army radar set, the SCR-268. The equipment, designed to direct searchlights to aircraft, was completed and demonstrated to the Secretary of War, Mr. Woodring, in May, 1937.

Meanwhile the British had embarked on a program of large scale "early-warning" radar development, which resulted in 1937 in the erection of the first CH (chain-home) stations on the Thames Estuary. These stations, operating at about 25 megacycles, were designed to detect aircraft flying over the British Channel. When Prime Minister Neville Chamberlain flew to Munich in September, 1938, these stations were put on 24-hr watch for the first time. Round-the-clock operation was resumed with the German occupation of Prague in November, 1938, and continued without interruption to the end of the War. The British development of radar during this period was well in advance of that in other countries, including America, and remained so until the pooling of British and American interests in 1940–1941.

Space does not permit a detailed account of the wartime employment of radar. Suffice it to say that from the Battle of Britain in 1940 (when radar control of 700 British fighter aircraft made it possible to defeat a force of some 2,000 German bombers) to the end of the war radar was credited with the leading role in many critical battles and naval engagements. Several Japanese warships were sunk in total darkness on the first salvo from American naval guns directed by radar. The Japanese were detected approaching Pearl Harbor on an SCR-270 early-warning radar. The coastlines of every belligerent country were protected by radar installations numbering several thousand. The total expenditure for radar research, development, and procurement in the period 1941–1945 was 2.7 billion dollars. Virtually the entire plant and personnel of the radio industry was turned over to radar work. Approximately 500 civilian scientists assembled by the National Defense Research Committee at the Radiation Laboratory, Massachusetts Institute of Technology, devoted the war years exclusively to the development of microwave radar (3,000 megacycles and higher frequencies). The Telecommunications Research Establishment, a laboratory of comparable size, was established under the Air Ministry in England.

All this activity was carried out with the greatest possible secrecy. Enemy activity in the field was at a high level but, par-

ticularly in airborne and shipborne equipment, was generally 6 months to a year behind Allied developments. The first public announcement of the existence of radar was made in June, 1941, when Lord Beaverbrook made an appeal for American technicians to man British radars, then called "radiolocators." A description of the radar principle was released in May, 1943. Since that time more detailed information has been released in technical publications.[1] But for the most part the record is kept in classified documents not widely available. For this reason references to past radar art cannot be quoted except in special instances.

For technical readers, the best index to wartime radar achievements is a statement of some of the magnitudes that have been attained, in power, frequency, accuracy of timing, and other essentials of radar system performance. The following is a brief summary:

Transmitted Power.—The term "megawatt" (one million watts) came into vogue in 1942 to describe the peak-power (pulse power) output of new tubes designed to provide the ultimate distance in radar detection. A form of cavity magnetron capable of developing peak power of 2 million watts at 10-cm wavelength was announced in that year. Similar power levels were achieved at lower frequencies with more conventional triode tubes. At higher frequencies (3-cm and 1-cm waves) peak power in the hundreds of kilowatts was available in cavity magnetrons. As is pointed out later in this chapter (Sec. 30), the peak power must be carefully distinguished from the average power generated, the average power being a better indication of the effectiveness of a radar signal in overcoming the effects of noise (that is, in detecting targets at great distances). Even when expressed in terms of average power, the figures are impressive, being roughly 1/1,000 to 1/100 the peak power, depending on the length of pulse generated and the pulse-repetition rate. One of the most astonishing achievements is the generation of 50 kw of power continuously at a wavelength of 50 cms; a feat accomplished in the resnatron[2] tube. These power levels exceed by several orders of magnitude those available immediately prior to 1940.

[1] See bibliography at end of this chapter.

[2] SALISBURY, W. W., The Resnatron, *Electronics*, **19**, (2) 92 (February, 1946).

Frequency.—British radar was first applied in the CH stations at 25 megacycles. The early U.S. Army developments that led to the SCR-268 began at 110 megacycles and the value was raised to 205 megacycles before production began. Later sets, widely used on naval vessels, were developed at frequencies approximating 600 to 1,500 megacycles. The microwave development that began in 1940 at frequencies near 3,000 megacycles steadily advanced in frequency to 10,000 megacycles and 30,000 megacycles. Development work on still higher frequencies is under way.

Timing Precision.—The generation of extremely short pulses and accurate timing of oscilloscope traces are necessary to achieve precision in distance measurement. Pulses under 5 μsec (microseconds) in length were the rule in virtually all equipment. Microwave (10-cm) equipment commonly used a 1-μsec pulse. For special purposes, 10,000-megacycle and 30,000-megacycle equipment used pulses as short as 0.2 μsec. The precision of timing was of the same order of magnitude. To measure distance to an accuracy within 10 yd, it is necessary to measure the radar reflection interval to 0.06 μsec. Circuits were developed to accomplish this task. Precision of timing over long intervals was exhibited by timing devices in the loran navigational system, which is closely allied to radar. Pulses were controlled in this equipment, accurate to a small fraction of a microsecond, over periods as long as 20 min., a timing precision of approximately one part in 10^9.

Radio-frequency Sensitivity.—The ultimate limit to receiver sensitivity is set by the noise level, which depends on the temperature and the bandwidth. At the lower frequencies, below 1,000 megacycles, receiver sensitivity approached within 5 db of the theoretical limit. At 3,000 megacycles the level was within 15 db of the theoretical limit and even at 30,000 megacycles the differential was reduced to 20 db. This is one of the most striking accomplishments of the entire development.

Apparatus.—The apparatus employed, particularly the r-f components and radiating structures, constitutes a whole new chapter in the radio art. The wave guide was exploited in every conceivable form. Beams were formed not only by reflectors, but by r-f lenses as well. Beams of r-f energy occupying a solid angle of one-half degree were produced at the highest frequencies.

The lower frequency components, at intermediate and video frequencies, made similar if less striking progress. Intermediate-frequency bands as wide as 20 megacycles were employed for special purposes, and i-f (intermediate-frequency) gains of 125 db were routine. The cathode-ray indicating apparatus was developed in some 16 separate and distinct modes of presentation of target coordinates.

A technical achievement of the greatest magnitude was the reduction of the experimental equipment to practical, serviceable gear that could stand severe battle conditions and be kept in condition by service personnel.

3. Methods of Radar Detection. *Pulse System.*—The essential components of a pulse radar system, shown in Fig. 1, are:

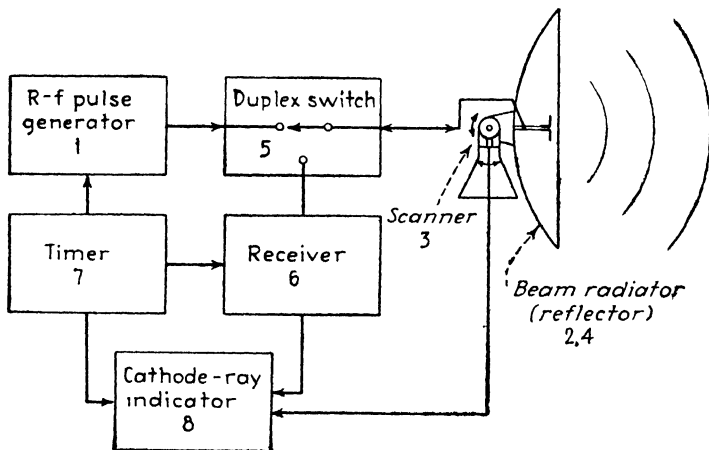



FIG. 1.—Essential components of a pulse radar system.

1. A source of short, high-power, h-f (high-frequency) pulses timed from a central source.

2. A beam-forming radiator that projects the succession of pulses into space along a beam of specified shape and energy distribution.

3. An electromechanical system capable of generating and transmitting voltages or currents which indicate the angular coordinates of the radiator (that is, the direction along which the beam is pointed).

4. A receiving antenna system the elements of which are usually common to the transmitting radiator.

5. A switching system ("duplexer" or "t-r (transmit-receive) box,") to prevent the transmitted signal from injuring the receiver and to prevent the transmitter circuits from absorbing the reflected signal.

6. A receiver of the highest possible sensitivity and appropriate bandwidth.

7. A timing circuit that measures the time of arrival of the reflected pulses relative to the time of emission of the transmitted pulses.

8. A cathode-ray indicator (or indicators) on which are presented the angular coordinates and/or distances of the targets within view of the radar.

The operation of these components is as follows: The transmitter is turned on for a very brief period (about 1 μsec) and is then turned off for a much longer period, several hundred or thousand microseconds. The transmitted pulse, directed by the radiator, encounters a target[1] in space. The signal induces a current in the target (the target need not be a good conductor if it is a good dielectric, since displacement current is sufficient), and the current causes reradiation from the target. The reradiated signal is propagated, generally, in all directions. A minute portion of the reradiated energy is propagated back to the receiver, the transmitter remaining idle during the reflection interval.

The received signal, after amplification and detection, is caused to modulate or deflect the beam of a cathode-ray tube. Other signals applied to the indicator tube indicate the duration of the reflection interval and the angular coordinates of the beam at the instant the reflected signal is received. The position of any target, which is not too distant or too small to reradiate a discernible signal, is thus indicated.

Some time after the reflected signal is received, the transmitter is turned on again for a brief interval and the process is repeated.

[1] The military origin of radar is clearly evident in this word, which has come to be the generic term for any reflecting object of particular interest in the path of a radar beam, whether or not it has military significance as a target. Other reflecting objects, not of particular interest and hence tending to obscure the target, are referred to as "background." The reflections from background are called "returns," that is, sea return or ground return, and the visual evidence of the background on the indicator screen is referred to as "clutter."

Any change in the distance or angular coordinates of the target is indicated by a corresponding change in the position of the target signal on the face of the cathode-ray indicator. If the target moves appreciably, it may be necessary to change the angular position of the beam in order to follow it, a process called "tracking." Alternatively, if the position of the target is not known, the beam may be swept methodically over a given region of space, searching for the target. If the beam traverses the region of space continuously and automatically, the process is known as "automatic search." If a limited region is searched by the operator, the term is "manual search."

FIG. 2.—Antenna of a ground-based radar for coastal gunfire control (type AN/MPG-1).

The operations just described are modified in practice in accordance with the particular function of the radar. A ground-based radar, set up to guard a coast against naval surface craft (Fig. 2), habitually searches over the seaward horizon, that is, the beam traverses in the horizontal direction. An airborne radar, searching for an enemy aircraft, searches in the forward direction and at the same time tilts the beam periodically above and below the horizontal (Fig. 15). A radar associated with an antiaircraft battery is designed to search skyward at all elevations up to the zenith and in all directions around the horizon. Moreover, a gun-pointing radar must indicate the angular coordinates of the target with extreme precision.

The pulsed system has proved to be amazingly flexible and adaptable to various needs. For this reason it has occupied nearly all the attention devoted to wartime development and

production, and occupies a correspondingly large portion of attention in this book.

The pulse system has several limitations, however, relative to the c-w system. The first and most basic limitation is the width of the band occupied in the ether spectrum, occasioned by the use of short pulses. This has not yet proved a serious practical

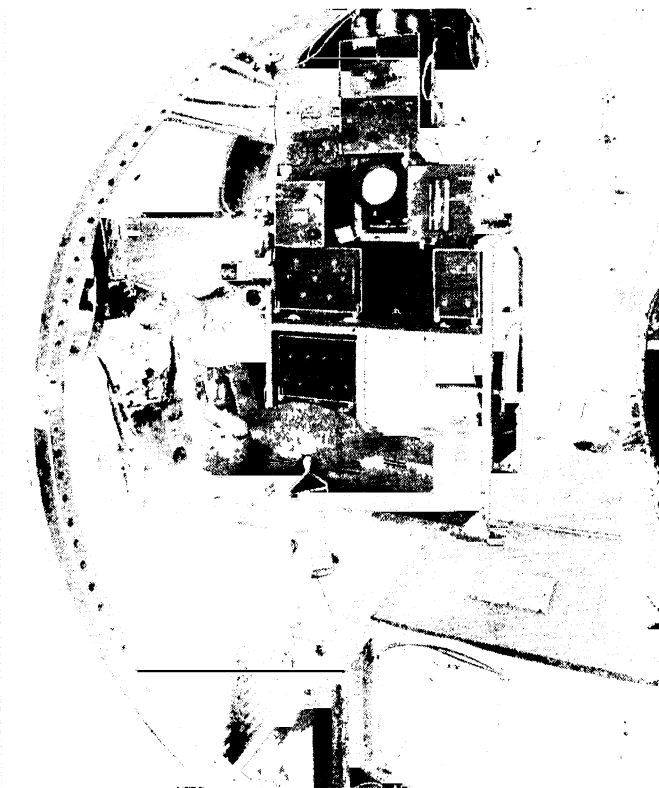



Fig. 3.—Components of an airborne radar for navigation and bombing (type AN/APQ-13).

limitation, since the new regions of the ether spectrum opened up by radar investigations have provided room for current needs. But as the uhf (ultrahigh-frequency) and shf (superhigh-frequency) regions are more thoroughly exploited, ether space may become scarce. The second limitation is that the pulse system does not make any clear distinction between moving targets and the stationary background against which they may be displayed.

Actually there is a distinction between moving and fixed targets in the pulsed system, but the distinction does not become clearly evident unless special means are taken (see Sec. 111) to enhance the effect. The unmodulated c-w system, on the other hand, is completely insensitive to fixed objects and gives a clear indication of moving targets against a fixed background

4. Continuous-wave System.—The c-w radar system, as the name implies, transmits continuously. A beam-forming radiator (Fig. 4) directs the signal to the target, which reradiates (scatters) the energy intercepted and returns a small fraction to the receiving antenna. Since transmitter and receiver are operating simultaneously and continuously, it is impractical to employ a common antenna system. Usually two similar structures are

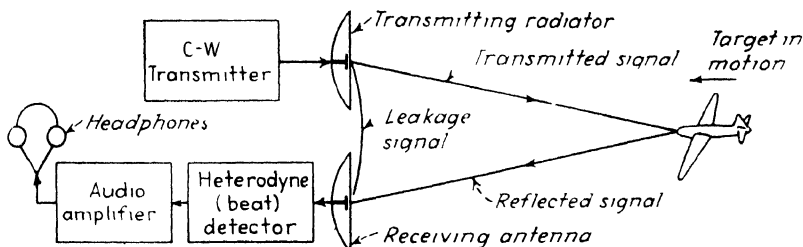



FIG. 4.—Essential components of a c-w radar system.

employed, side by side, and so oriented that only a small fraction of the transmitted power leaks directly to the receiver. Distinction between the transmitted signal and reflected energy arriving from a target is made on the basis of slight changes in the carrier frequency.

In the unmodulated c-w system the necessary shift in frequency is caused by motion of the target relative to the radar. The shift in frequency, known as the "Doppler shift," amounts to 30 cps per 100-megacycle carrier frequency, for each 100 mph relative velocity. The frequency shift is ordinarily in the audio frequency range and can be detected aurally by mixing the reflected signal and the leakage signal in a detector that develops the Doppler shift as an audio output. The Doppler-shift frequency can be applied to a cathode-ray indicator that also presents information on the orientation of the transmitted beam.

The difficulty with this system is that it does not reveal the distance of the target, nor is it able to distinguish one target from

another when both happen to lie in the same direction from the radar. Moreover, it is blind to stationary or slowly moving targets. The advantages are the converse of the pulse system limitations: the ether space required is as small as the frequency stability of the transmitter and receiver will permit, and the system is insensitive to stationary background reflections.

A variation of the c-w radar system is the f-m (frequency-modulated) version. In this version, the transmitted signal is swept in frequency over a small range about the carrier value.

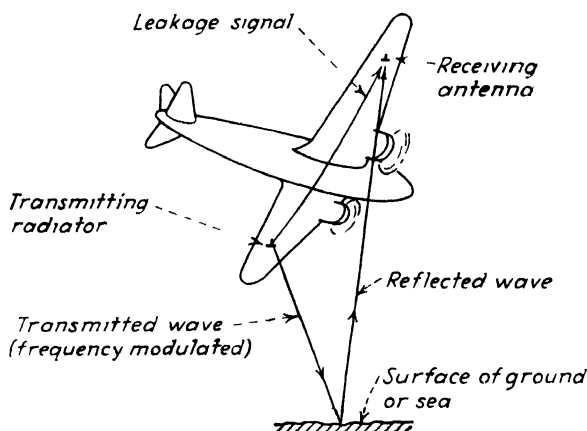



FIG. 5.—Use of c-w radar to determine height of aircraft above terrain.

The reflected energy preserves the frequency with which it left the transmitter and returns to the receiver only after the transmitter frequency has changed by an appreciable amount. Thus the reflected signal and the leakage signal are separated in frequency by an amount that can be detected as a beat frequency and indicated aurally or otherwise. The system operates equally well on fixed and moving targets (except in the improbable event that the Doppler shift due to relative velocity of the target cancels the shift due to the frequency modulation). The system indicates the distance to a fixed target, since the frequency difference introduced is proportional to the time of travel of the reflected energy, and the beat note may be fed to a frequency meter whose indicator is calibrated directly in distance units. The distance measurement is subject to error if the target is moving. The f-m radar has found its widest use as an absolute

altimeter[1] for aircraft (Fig. 5), the ground directly under the aircraft serving as the reflecting surface.

5. Technical Specifications of a Pulse Radar System.—The performance of a radar system involves the detection of a given target at a given distance with a given accuracy of position. Corresponding to these performance specifications are technical specifications that may be classed in four categories: the pulse specifications, the r-f (radio-frequency) specifications, the scanner specifications, and the indicator specifications. The pulse specifications determine the range of distances over which the equipment is operative. The r-f specifications relate to the precision of measuring the angular coordinates of the target and also determine to a large extent the maximum distance at which a given target can be detected. The scanner and indicator specifications relate to the final use of the target information, since the scanner finds and the indicator presents the target coordinates in a form appropriate to the use of the equipment.

6. Pulse Specifications.—A radar pulse is described in terms of its *peak power*, its *width* (duration of the pulse, sometimes called

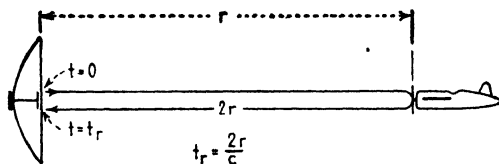



FIG. 6.—Relationship between time of travel and distance to the target.

"pulse length"), and the *pulse-repetition rate* (number of pulses per second). The peak power, among other factors, determines the maximum distance at which the system will detect a target. The length of the pulse determines the minimum range of the system and has an important effect on the precision with which the target distance can be specified. The pulse-repetition rate limits the maximum range of the system and determines the speed with which a given region in space can be scanned in searching for a target.

To make clear the manner in which the pulse specifications govern these aspects of radar performance, it is necessary to examine the numerical values associated with the velocity of radio propagation (Fig. 6). Within the accuracy demanded of

[1] FINK, D. G., F-M Radar Altimeter, *Electronics*, **19** (4), 130 (April, 1946).

radar equipment the velocity of radio propagation may be taken as the velocity of light, 2.998 × 10^8 m per sec, or 299.8 m per μsec. Table I gives the velocity expressed in units used in radar work.

TABLE I

Velocity	Reciprocal Velocity
299.8 m/μsec	0.003336 μsec/m
983.6 ft/μsec	0.001017 μsec/ft
327.9 yd/μsec	0.003050 μsec/yd
0.1863 statute miles/μsec	5.368 μsec/statute mile
0.1618 nautical miles/μsec	6.180 μsecs/nautical mile

The reciprocal velocities show the fraction of a μsec consumed by the radio wave in traveling one meter, foot, yard, etc.

In radar work, principal interest is attached to the reflection interval, that is, the length of time it takes a pulse to go to the target and return to the source. The distance traveled by the wave is twice the distance to the target, and the reflection intervals are accordingly twice the reciprocal velocities given in Table I. By the same token, the distance to the target per unit time interval is one-half the distance traveled by the wave in the same time, that is, one-half the velocity figures of Table I. These values are shown in Table II.

TABLE II

Radar Distances	Reflection Intervals
149.9 m/μsec	0 006671 μsec/m
491.8 ft/μsec	0.002033 μsec/ft
163.9 yd/μsec	0.006101 μsec/yd
0.0932 statute miles/μsec	10.735 μsecs/statute mile
0.0809 nautical miles/μsec	12.361 μsecs/nautical mile

The pulse width is limited by the fact that the transmitter must be turned off in time to permit the receiver to perceive reflections from the nearest target of interest. Stated differently, if the pulse width exceeds the reflection interval of the nearest target, the reflected signal cannot be distinguished from the transmitted signal. The numerical relation is given in Table II. If it is desired to detect targets at distances as short as 492 ft, the pulse must not be wider than 1 μsec. Actually the minimum detection distance of a 1-μsec system is somewhat greater than 492 ft, since additional time is required to permit the receiver system to recover from the effects of the transmitted signal.

A narrow pulse is also of assistance in revealing a small target

against a reflecting background.) To understand the effect of pulse width in this problem, it must be appreciated that each pulse has a definite physical length in space, equal to the distance propagated while the pulse is leaving the radiator (see Fig. 7). A 1-μsec pulse extends some 984 ft in space, and is thus capable of illuminating the region in front of and behind a target whose dimensions are smaller than this. The narrower the pulse, the more energy is directed, proportionately, to the target and the less energy to the reflecting background.

Finally, a narrow pulse is essential in any system that must specify the distance of a target with precision. The precision with which a pulse may be timed, as it leaves the transmitter and

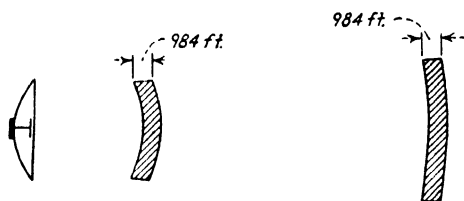



FIG. 7.—Each pulse extends in space an amount equal to the distance traveled during emission from the radiator (one μsec pulse shown).

arrives back at the receiver, is directly proportional to its width, other factors, including pulse shape, remaining unchanged.

By similar reasoning it may be shown that the interval between successive pulses corresponds to the maximum detection distance of the system. For it is clear that the transmitter, having been turned off to permit perception of echoes from near-by targets, must remain inactive so long as echoes from more distant targets are to be received. Again the numerical relation is given in Table II. If targets are to be received from a maximum distance of say 93.2 miles, the transmitter must remain off the air for a period of at least 1,000 μsec. In the numerical example given (1-μsec pulses spaced 1,000 μsec) the radar will detect targets from approximately 500 ft to approximately 90 miles, provided that it has sufficient signal energy and receiver sensitivity to reach the outer limit.

Several quantities derived from the pulse width and interval between pulses are of practical convenience. These quantities are the pulse-repetition rate (pulse rate) and the duty cycle. The pulse rate, the number of pulses transmitted per second, is the

inverse of the total period of the pulse waveform, as shown in Fig. 9. Suppose that the pulse width is 1 μsec and the inactive period between pulses, in close similarity to the preceding example, is 999 μsec. Then the period of the pulse waveform is 1,000 μsec, or 0.001 sec, and the pulse rate is 1,000 per second.

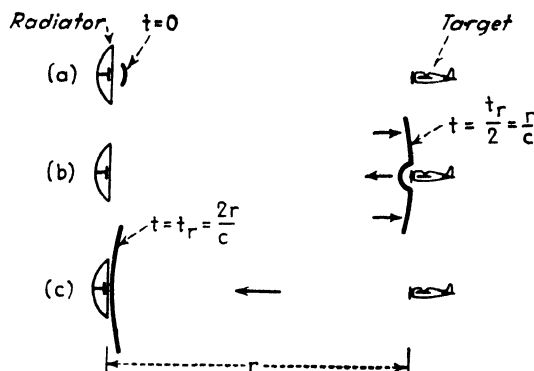



FIG. 8.—Three stages in the transmission and reflection of a radar pulse.

The duty cycle of a pulse waveform, as its name suggests, is the ratio of the pulse width to the interval between like portions of successive pulses. In the example just quoted (one microsecond pulse every 1,000 μsec) the duty cycle is 1/1,000 = 0.001. The

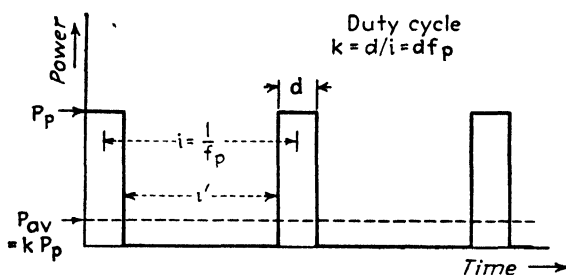



FIG. 9.—Basic geometry of the rectangular pulse, showing average and peak values.

duty cycle is a convenient index to the average power transmitted, since (if the pulses are rectangular) the average power is equal to the peak power times the duty cycle. In the case of a non-rectangular pulse, the duty cycle is computed from the area under the pulse. Figures 9 and 10 show the relationship between

pulse width, repetition rate, pulse interval, duty cycle, and peak and average power. The quantities involved are

Peak power $= P_p$ watts
Pulse width (duration) $= d$ sec
Pulse interval $= i$ sec
Pulse-repetition rate $= f_p = 1/i$ cps
Duty cycle $= k = d/i = df_p$
Average power $= P_{av} = kP_p = df_pP_p$ watts
Inactive interval $= i' = i - d$

The pulse-repetition rate, being the inverse of the pulse interval, determines the maximum range of detection. The pulse rate is also of great importance in determining how fast a radar beam

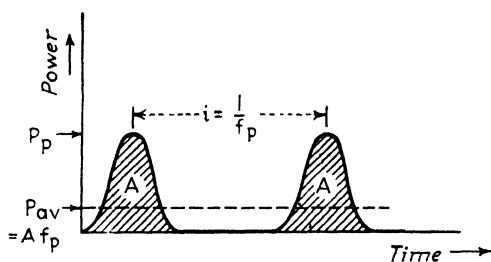



FIG. 10.—Nonrectangular pulse. The average value is measured in terms of the area A under each pulse.

may be swept through space. If the pulses are sent at a slow rate and if the beam is narrow and rapidly moving, it is possible to miss the target by passing over it in the interval between pulses. To guard against this contingency the pulse rate must be coordinated with the beam width and angular scanning speed, as discussed in detail later in this chapter (Sec. 9).

The peak power of the pulse is obviously a matter of primary importance since the greater the power the greater the distance at which targets may be perceived. The quantitative relationship between peak power, target distance, and other pertinent factors is known as the "radar equation." Its derivation is given in Sec. 28, in this chapter.

7. Radio-frequency Specifications.—The choice of the operating wavelength of a radar system depends fundamentally on many considerations. Only the principal factors are discussed in this introduction. First and foremost is the fact that the wavelength employed must be short compared with the dimen-

sions of the radiator if a narrow beam is required This implies frequencies of the order of 100 to 300 megacycles in fixed ground installations where radiator dimensions in the tens of feet can be allowed (Fig 11) It implies frequencies of the order of 3,000

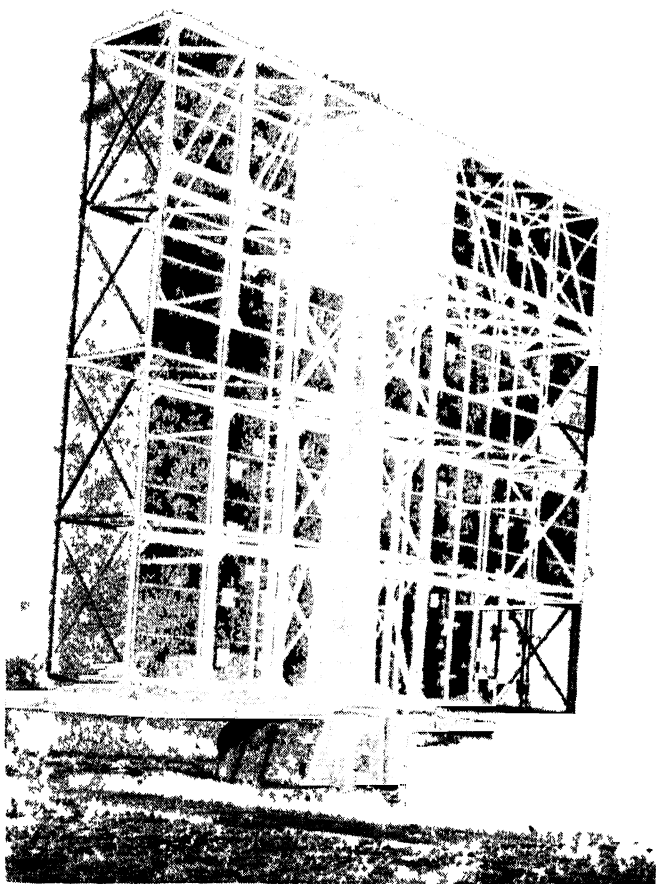



FIG 11—Typical "mattress" array of dipoles (British fighter-director radar)

megacycles or higher in airborne equipment where similar directivity must be achieved in radiators measured in the tens of inches (Fig 12)

It follows then that high frequencies must be used when precision is required in the determination of the angular coordinates of the target, if the radiator must be of convenient size

Two additional considerations in the choice of operating frequency are transmitter power capability and receiver sensitivity The quantities are generally low at the higher frequencies, and thus the maximum detection distance of the h-f equipment tends to suffer in comparison with the lower frequency equipment, other factors being equal)

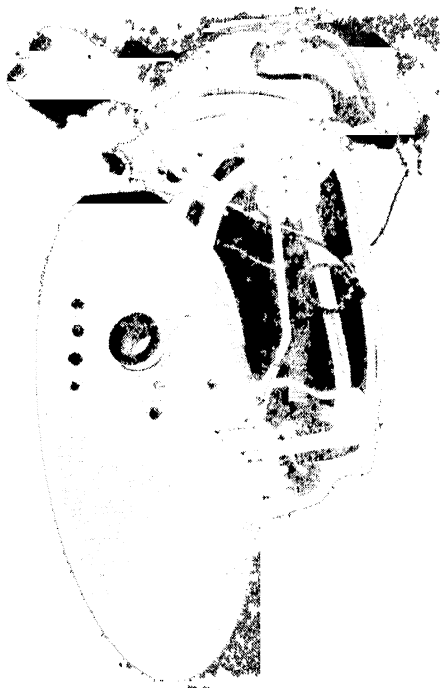



FIG. 12.—Typical airborne antenna, a 3-cm unit used for navigation and bombing (type AN/APS-15)

A fourth consideration is the effect of ground reflections on propagation. In l-f (low-frequency) equipment, intended for searching in a horizontal direction, it is virtually impossible to avoid directing some of the energy toward the ground or water surface. Reflections from this surface combine with the directly propagated waves to form an interference pattern. This pattern displays discrete lobes of signal, the horizontal angle of the lowest lobe depending inversely on the height of the transmitting radiator above the earth, in wavelengths. By using higher frequencies, it is possible to produce a narrower beam and hence

restrict the radiation toward the surface. High frequencies are thus indicated for the detection of low-flying aircraft or surface craft.

When the target reradiates the radar signal, it is clear that the wavelength, relative to the target size, has an important effect on the strength of the reflection. This effect is very pronounced when the target size approaches a half wavelength. Moreover the reradiating ability of a target is much reduced when the target is much smaller than a quarter wavelength. This factor indicates the use of high frequencies for small targets, such as projectiles in flight.

In military applications the choice of wavelength is also guided by knowledge of enemy activity. It is advantageous to use h-f equipment, on the assumption that it will be more difficult for the enemy to interfere with or detect such transmissions.

The choice of the wavelength, with a radiator of given area, determines the power gain of the radiator, and this power gain applies directly to the peak power of the pulse. The connection between pulse specifications and operating frequency is revealed in further detail in the radar equation, Sec. 28.

Aside from the choice of the operating frequency, the r-f specifications have to do primarily with the particular equipment employed, such as the type of oscillator employed, magnetron or triode, the form and size of the radiator and reflector, the range over which the equipment may be tuned, the efficiency of the duplex switching system, etc.

One other r-f specification that has quite general significance is the receiver noise figure. This figure represents the circuit noise present in the output of the receiver, measured in decibels above the theoretical lower limit of noise $kT\,\Delta f$, where T is the absolute temperature, Δf is the nominal bandwidth of the receiver system, and k is Boltzmann's constant. This figure, it should be noted, differs from the noise power that is generated in a resistor, $4kT\,\Delta f$. The derivation of the $kT\,\Delta f$ quantity is given in Chap. II, Sec. 60.

As previously mentioned the noise figures achieved in practical radar receivers vary from 5 db to 20 db above the theoretical limit, the higher figures being generally associated with the higher frequencies.

Another item of interest in the r-f specifications is the fre-

quency stability achieved in transmitter and receiver circuits. This characteristic determines the additional bandwidth, and consequent additional noise, that must be allowed for frequency drift above the normal bandpass requirements.

8. Scanner Specifications.—The word "scan" has been restricted in radar work to denote the motion of a r-f beam through space in searching for a target, and it does not refer to the deflection of cathode-ray beams. The scanner is the mechanical and electrical structure that rotates or otherwise moves the radiator so that it points in successively different directions.

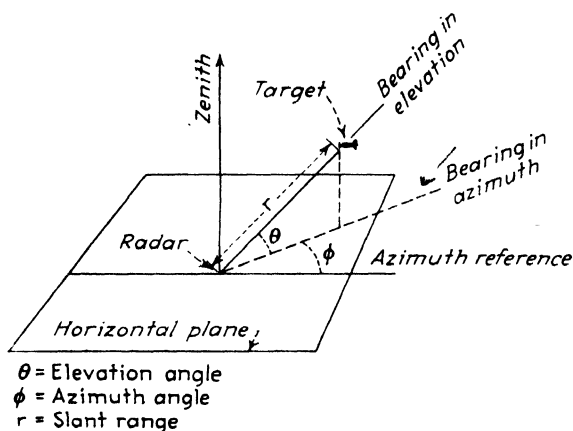



FIG. 13.—Spherical coordinates used in specifying target position relative to radar.

The types of scanning are denoted by the path described in space by a point on the radar beam, for example, circular, helical, conical, and spiral. There are many variations, such as the sector scan, which covers a restricted portion of a circular scan. Certain high-speed scanners describe a back-and-forth motion that resembles the scanning motion in a television picture and hence is called "television scan."

Before going into the details of these scanning motions it is necessary to establish the terminology of target coordinates that has been adopted as standard in radar work. Following established military and naval practice, the target position is specified in spherical coordinates with the radar equipment at the origin, as shown in Fig. 13. The three coordinates are *range* (r), *azimuth* (ϕ), and *elevation* (θ). The range (strictly speaking

"slant range") is the length of the radius vector from the origin (the radar) to the target. The azimuth is the angle, measured in a horizontal plane, between a reference direction, for example, north and the projection of the radius vector on the horizontal plane. The elevation (sometimes called "angular height") is the angle measured in a vertical plane between the radius vector and the horizontal plane. A scanner scans sidewise in the azimuth angle and up and down in the elevation angle. The range is, simply, the distance to the target.

Other important target coordinates that may be used in particular instances are *altitude* or *height* (the vertical distance from the target to the horizontal plane) and the *horizontal range* (or *ground range*), the length of the projection of the radius vector on the horizontal plane.

9. Circular Scanning.—The circular scanner describes a path in azimuth only, viewing the horizon and turning through an

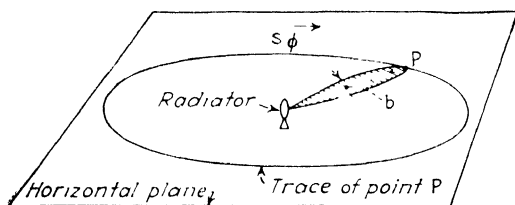



FIG. 14.—Geometry of circular scanning.

arc from a few degrees (sector scan) to 360 deg. A single quantity is sufficient to specify a circular scan, the angular rate of rotation s_ϕ radians per sec. As has been mentioned, this quantity must be coordinated with the rate of transmitting pulses to be sure that each point receives at least one pulse as the beam passes over it. Assume that the effective angular width of the beam is b radians and that it rotates s_ϕ radians per sec. Then a given point in the plane of rotation remains in the beam for b/s_ϕ sec. At least one pulse and preferably a greater number of pulses n (*e.g.*, $n = 5$) must be radiated in that time to assure that n pulses reach each point in the plane of rotation. Hence the required pulse-repetition rate is

$$f_p = \frac{ns_\phi}{b} \quad \text{pulses per sec} \tag{1}$$

The angular spacing in azimuth, between pulses, is

$$\phi_p = \frac{s_\phi}{f_p} \quad \text{radians} \tag{2}$$

Circular scanning is used in ground-based radars to search the horizon for surface craft, low-flying aircraft, or aircraft at a distance. It is used in airborne radars to search the area about the aircraft for other aircraft at nearly the same altitude, or to search the ground for targets and terrain features. Figure 14 shows the parameters that characterize circular scanning.

// **10. Helical Scanning.**—In helical scanning a point on the radar beam traces out a distorted helix in space, as shown in Fig. 15. The motion is a combination of circular scanning in azimuth at an angular rate s_ϕ radians per sec with a slower vertical tilt in eleva-

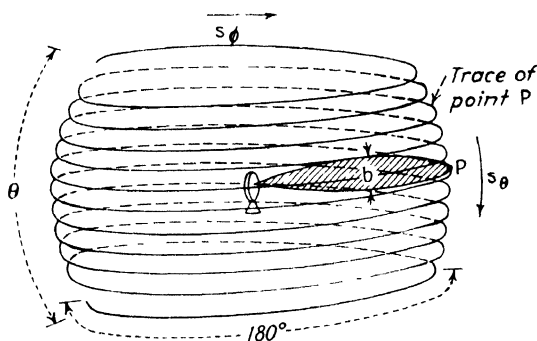



FIG. 15. Geometry of helical scanning.

tion at a rate s_θ radians per sec.) This type of scanning is used in airborne equipment for the interception of enemy aircraft by searching the space above and below the horizontal plane. In a typical example, the elevation angle scanned varies from 10 deg below the horizontal to 60 deg above, reflecting the greater interest on the part of the pilot in planes attacking from above. Helical scanning is also used in ground-based and shipborne radars to search the sky for aircraft throughout the full range of azimuth and elevation angles from the horizon upward to the zenith.

The specification of a helical scan is evidently more complex than that of the circular scan. Not only is it essential to coordinate the pulse rate with the azimuth scanning rate, but also it is necessary to ensure that successive turns of the helical motion overlap one another, so that no portion of the space searched is

left uncovered. In practice a twofold overlap is considered ample, that is, the beam must pass over a given point in space twice on successive turns of the helix. Thus the permissible motion in elevation, per turn of the helix, is $b/2$ radians. Since a single turn of the helix is transversed in $2\pi/s_\phi$ sec, it follows that

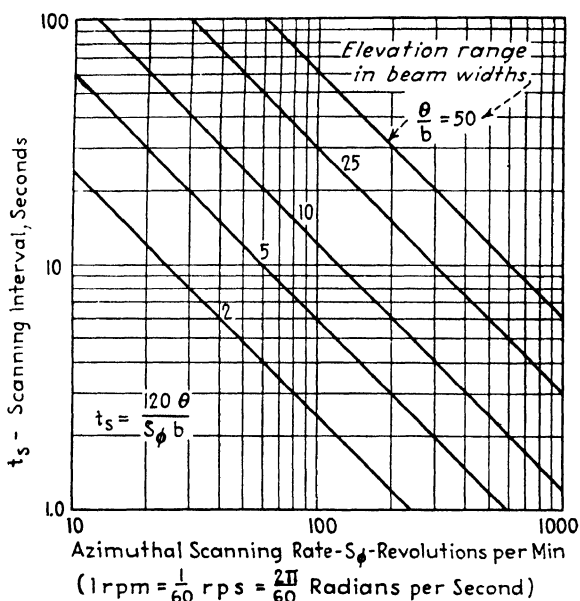



FIG. 16.—Relationships between scanning rate, beam width, elevation range and scanning interval.

the rate of scanning in elevation is

$$s_\theta = \frac{bs_\phi}{4\pi} \qquad \text{radians per sec} \tag{3}$$

If the total range of elevation to be covered is θ radians, the total time required to scan the region of space is

$$t_s = \frac{4\pi\theta}{bs_\phi} \qquad \text{sec} \tag{4}$$

This is the so-called "scanning interval." Figure 16 is a plot of Eq. (4).

Equation (4) is the clue to one of the principal limitations of radar scanners, that is, the time required to cover a given region in space. To cite a typical instance, a 3,000-megacycle radar

with a 30-in. paraboloid reflector has a nominal beam width b of 8 deg. As applied in aircraft interception work the beam scans a region $\theta = 70$ deg in elevation and scans in azimuth at a rate s_ϕ of 3 rps, or 6π radians per sec. It follows that the scanning interval (time required to complete the scanning motion) is

$$t_s = \frac{4\pi \times 70}{8 \times 6\pi} = 5.8 \qquad \text{sec}$$

It thus appears that it is difficult to repeat an indication of a given target more often than once in several seconds, and it follows that the indicator device must store the information over several seconds to display target motion continuously. Since cathode-ray phosphors possessing the necessary persistence are available this is not a serious limitation. More serious is the fact that the target may move out of the field of view during the scanning interval. This may well happen if the target is a fast aircraft near the radar. Hence fast scanners are required to follow fast-moving targets.

11. Conical Scanning.—In conical scanning, a point on the radar beam describes a circle at the base of a cone, and the axis

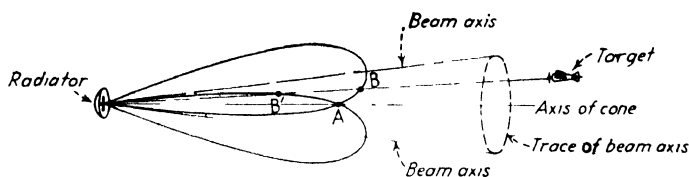



FIG. 17.—Geometry of conical scanning.

of the beam is the generatrix of the cone. Conical scanning, shown in Fig. 17, is used to determine the azimuth and elevation of a target to a precision considerably greater than the width of the beam itself. It is used in radars to direct the fire of anti-aircraft guns and in similar high-precision applications.

The quantity specifying the conical scan is the angular scanning rate about the axis of the cone in radians per second. Equation (1), derived for circular scanning, also applies to conical scanning, if the quantity s_ϕ is taken as the angular rate of scanning about the axis of the cone.

The principle of conical scanning is as follows: If a target lies within the cone but off the axis of the cone, it receives different amounts of energy from the beam as the beam moves over the

surface of the cone. The variation in incident energy becomes less as the target approaches the axis of the cone, and when the target is on the axis the power delivered does not vary. By detecting the variation in amplitude of the received pulses it is possible not only to point the axis of the cone directly at the target, but also to perceive by what angles, in azimuth and elevation, the target direction departs from the axis of the cone. An automatic target-following or "tracking" system is thus made possible. Since the rate of change of power with angle is great at the point A (see Fig. 17), it follows that the direction along which the signal variation ceases is sharply defined. Angular precision of a few hundredths of degree is attainable with this method.

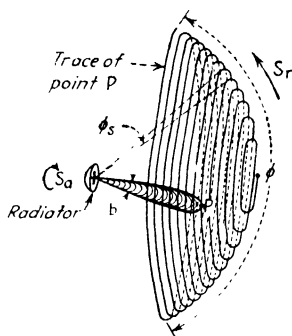



FIG. 18.—Geometry of spiral scanning.

Conical scanning is produced, mechanically, by rotating or nutating a *dipole antenna in a small circle about* the focus of a paraboloid reflector. Details are given in Chap. IX.

12. Spiral Scanning.—Spiral scanning is a generalized form of the conical scan, which covers a region in space, in azimuth, and in elevation, for searching purposes. It is often used in connection with gun-laying radars as an auxiliary searching method to find the target, after which conical scanning is used to follow the target automatically.

The geometry of the spiral scanner is correspondingly more complex than that of the conical scanner. Figure 18 shows the pertinent quantities that describe the motion. Let ϕ be the total range covered in elevation or azimuth, measured at right angles to the axis of the spiral, s_a be the angular rate of rotation along the spiral, and s_r the angular rate at which the beam moves, from center of spiral outward, at right angles to the turns of the spiral. Then, to assure a twofold overlap between successive turns of the spiral,

$$s_r = \frac{bs_a}{4\pi} \quad \text{radians per sec} \tag{5}$$

which has the same form as Eq. (3) for helical scanning.

To assure that at least n pulses fall on each point of space as the

beam moves over it, attention must be paid to the outermost turn of the spiral since the space traveled by the beam between pulses is greatest along this line. Here the rate of pulses transmitted must be

$$f_p = \frac{ns_a}{b} \sin \phi/2 \qquad \text{pulses per sec} \tag{6}$$

If the pulse rate is made adequate for the outer turn of the spiral, it is adequate for all the inner turns. In fact, since the inner turns are completed in less time than the outer, the number of pulses that fall on a target near the center of the pattern is larger in proportion, provided the pulse rate is fixed. To avoid this unequal distribution of pulses, a variable pulse rate is sometimes used, such that the number of pulses delivered to each point in space is constant throughout the scanning pattern.

The spiral scanner is outstanding in its ability to cover a region of space in a very short time, since it is not necessary to move the paraboloid reflector. The scanning pattern shown in Fig. 18 is covered in less than a second in typical equipment.

13. Lobe Switching.—When it is necessary to determine angular coordinates with high precision using a l-f radar, the

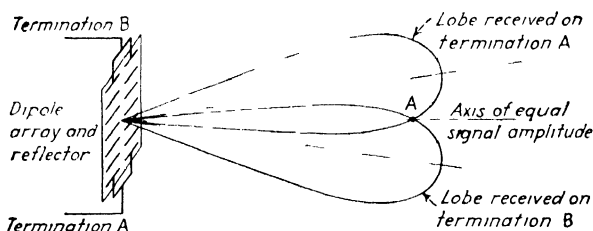



FIG. 19.—Lobe switching—a method of refining angular measurements.

large size and weight of the dipole array make it impractical to employ conical scanning. Instead a similar technique, called "lobe switching," is used. As is shown at length in Sec. 94, Chap. IV, the direction of the beam created by a dipole array can be shifted by adjusting the relative phases of the currents in the various dipoles. By setting up two phase-adjusting circuits and connecting them alternately, the direction of the beam may be shifted back and forth periodically. In practice the beam, or "lobe," is switched back and forth several times per second over an angle equal to a fraction of the beam width. The strength of the signal as it reaches the target and the strength

of the subsequently received reflections vary as the lobe is switched, unless the target lies in the plane bisecting the angle formed by the two lobe positions (see Fig. 19). The array is oriented until the strength of both sets of pulses is the same, as revealed on a cathode-ray indicator whose trace is switched synchronously with the lobe. Angular precision of the order of one-half degree is attainable using this method at carrier frequencies of the order of 200 megacycles.

14. The Indicator Specifications.—The radar indicator is the device that presents information concerning the presence and position of the target. In pulse radars, the indicator is a *cathode-ray tube whose beam is modulated in intensity and/or deflected in position to indicate the target coordinates.* Intensity *modulation,* *turning the* beam on and off, is generally reserved for indicating the presence of a target and distinguishing it from other targets. This leaves the three target coordinates (r, ϕ, and θ) to be indicated by two coordinates of deflection,[1] the x and y directions. It is thus generally impracticable to represent all the target information on the face of a single cathode-ray tube, unless special devices are used. Generally two and sometimes three or more cathode-ray indicators are necessary to provide all the target information.

Sixteen different arrangements of coordinates have found some degree of use in radar indicators, but four types account for the majority of practical applications. These are classified as the type A, type B, type C, and ppi (type P) presentations.

15. Type A Indicator.—The type A indicator ("A scope"), the earliest and most generally used form of radar display, closely resembles the linear-sweep deflection of a standard cathode-ray oscilloscope. The cathode-ray beam is swept at uniform velocity from left to right, creating a uniform time scale (and hence uniform distance scale) in the horizontal direction. The target is indicated by applying the demodulated echo signal to the vertical deflection plates of the indicator tube. The target signal may deflect the beam upward or downward, although the upward direction is more commonly used. The type A indicator, shown

[1] The term "scanning," commonly used in cathode-ray technology to describe the motion of the luminous spot across the screen, is reserved in this book to describe the motion of the r-f beam in space. The term "deflection" will be used exclusively to describe the motion of cathode-ray beams.

in Fig. 20, reveals the presence of the target and its range, but gives no angular (azimuth or elevation) information. The zero of the time base is set by the transmitted pulse, which is displayed, at least in part, at the left-hand edge of the trace.

The type A indicator is used to indicate the presence and range of a target, but, since it does not reveal angular coordinates, its practical utility is somewhat limited. The universal presence of the type A indicator is explained by the need of measuring the strength of the received echo. Such signal strength measurements are essential in adjusting (for example, tuning) the various elements of the radar to the point of maximum sensitivity. The type A indicator is also very helpful in revealing the nature of the

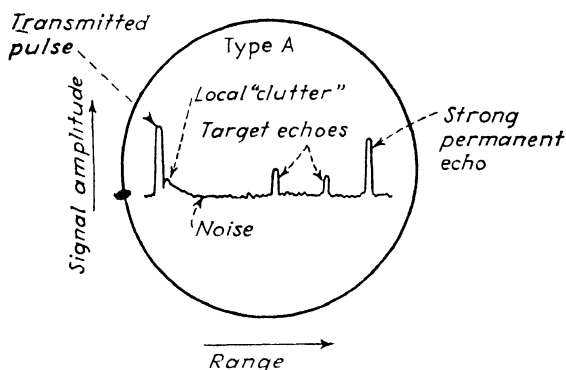



FIG. 20.—Type A indicator.

target, since the echo is displayed in two dimensions, time and voltage, whereas in other indicators the echo is displayed merely as a change in brilliance of the spot cathode ray. This aspect of indicator performance is discussed more fully in Sec. 25, in connection with the types of radar targets.

16. Type B Indicator.—The type B indicator ("B scope"), shown in Fig. 21, presents the target by intensity modulation of the cathode-ray beam. The position of the resulting spot of light indicates the range and azimuth of the target, in rectangular coordinates. The range is displayed vertically and the azimuth horizontally. This type of indicator is particularly well suited to radars that employ circular scanning over a limited sector, since the scanner then searches for targets in range and azimuth only, and the limited azimuth sector can be readily displayed in rectangular coordinates. The transformation from the polar

coordinates of the scanner to the rectangular coordinates of the indicator is illustrated in Fig. 22. It is evident that the distortion consists of spreading out the angle scale at close ranges.

The type B deflection is formed by deflecting the beam vertically by a linear sweep circuit, starting at the bottom at the

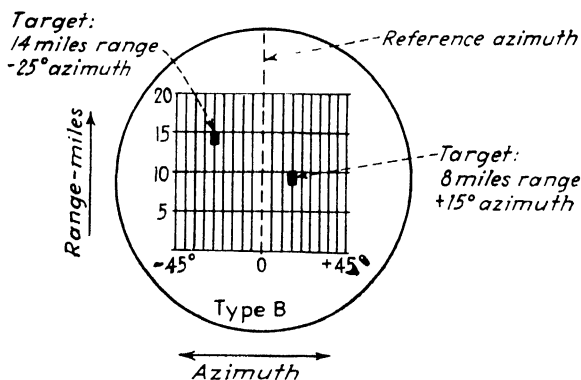



FIG. 21.—Type B indicator.

instant each pulse is transmitted. The demodulated reflected pulses are applied to the control electrode of the cathode-ray tube and brighten the spot, provided the echo voltage is strong enough to be discerned above the noise. The horizontal deflec-

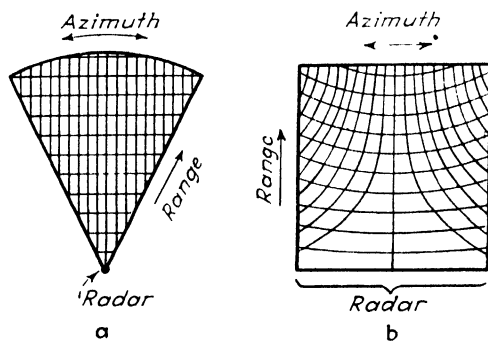



FIG. 22.—Distortion (b) introduced by type B indicator relative to space scanned (a).

tion is obtained from a potentiometer mounted on the scanner that feeds to the horizontal deflection coil a current proportional to the angular rotation in azimuth. In general the type B indicator gives a distorted view of the contents of the space in front of the radar, projected onto a horizontal plane.

17. Type C Indicator.—The type C indicator ("C scope") indicates the presence of the target by intensity modulation and indicates the azimuth and elevation of the target in rectangular coordinates, as shown in Fig. 23. No range information is displayed. The type C indicator is used usually in conjunction with a type A or type B scope to give the target coordinates in

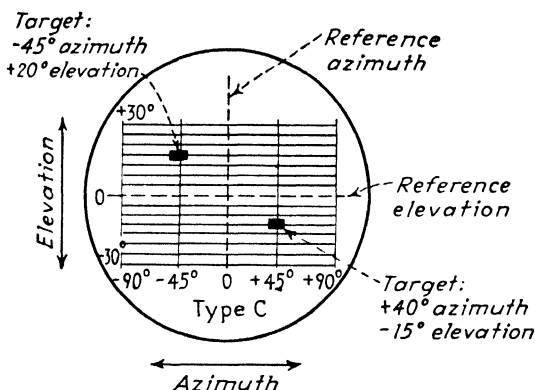



FIG. 23.—Type C indicator.

three dimensions. The type C indicator suffers from an inherently poor signal-to-noise ratio, since all the noise received throughout the pulse repetition interval is imposed in the luminous spot, whereas in the type B scope the noise is spread out along each vertical line and the signal has to compete only with the noise present at the instant the signal is received. This

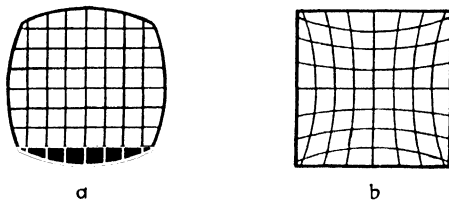



FIG. 24.—Distortion (*b*) introduced by type C indicator, relative to space scanned (*a*).

limitation of the type C indicator may be removed at some sacrifice in scanning speed by cutting off the receiver except during a particular interval, or gate period. The position of this gate is moved outward, searching in range for targets.

The transformation from the polar coordinates of the scanner to the rectangular coordinates of the type C indicator is shown in

Fig. 24. The effect is that of projecting the contents of the space scanned onto a vertical plane at right angles to the direction of view. The indicator is useful in searching for targets in three-dimensional space, particularly in airborne radar using helical scanning. When the target is detected, information on its range is secured by reference to a type B indicator (or, more rarely, a type A indicator).

18. Plan-position Indicator.—The plan-position indicator (ppi or Type P), is one of the most useful radar presentations because

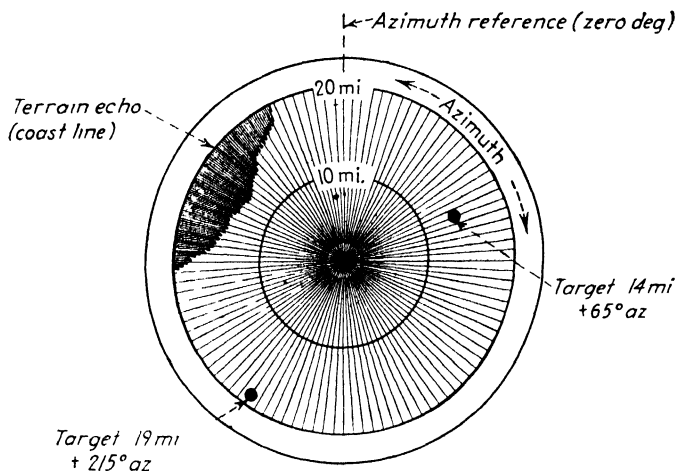



FIG 25.—Plan position indicator (type P or ppi).

it gives a maplike presentation throughout 360 deg, or over any smaller sector of particular interest. As shown in Fig. 25, the ppi imitates the scanner by employing polar coordinates, the range of the target being indicated by radial distance from the center of the pattern and its azimuth being indicated as an angle relative to a fixed reference. When employed in ground-based radars, the ppi indication is directly comparable with a map, the center of the presentation corresponding to the position of the radar on the map. When employed in airborne radar, for precise work it is necessary to convert the slant range indicated on the ppi to horizontal range as a function of altitude, but otherwise the correspondence to a map applies.

The ppi is closely related to the type B indicator, since both present range and azimuth information and both may be used with the same types of scanners. The ppi has a great advantage

in that it can cover azimuth angles up to a complete circle without distortion, whereas the type B indicator gives excessive distortions when the sector exceeds 45 deg. The ppi may be used with other indicators when the scanning covers azimuth and elevation, but generally the ppi finds its greatest usefulness in horizontal search scanning, that is, in azimuth only.

The ppi is not an easy presentation to produce electrically, and it did not appear widely in radar work until the type A, B, and C indicators had become well established. Details of ppi deflection systems are discussed in Chap. XI. One simple form of ppi

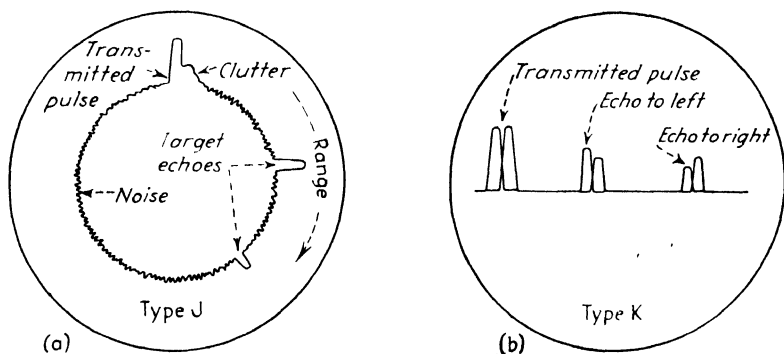



FIG. 26.—Modifications of the type A indicator.

system is a magnetic deflection yoke, carrying a sawtooth deflection current that displaces the luminous spot outward from the center of the tube at a constant rate, the motion starting with the initiation of each transmitted pulse. The spot is brightened by the application of the target echo signal. The yoke is rotated bodily about the neck of the cathode-ray tube by a gear drive synchronized with the azimuth motion of the scanner.

19. Modifications of the Type A Indicator.—Type J, K, L, M, and N indicators, shown in Figs. 26 and 27, are modifications of the "range-only" deflections of the type A indicator. Type J is a circular sweep of constant speed with radial deflection to indicate the signal. It is most simply produced in a tube having a radial deflection electrode. It is used in accurate range-finding applications, the circular sweep serving as an expanded portion of the range coordinate. Its principal advantage is the increased length of sweep, relative to a linear deflection, for a given diameter of cathode-ray tube.

Type K is a modified A scope used to indicate the relative strength of the two sets of pulses received in the lobe-switching system. Two sweeps are superimposed, but are slightly displaced along the length of the sweep, so that the two sets of pulses are displayed side by side. It is then comparatively simple

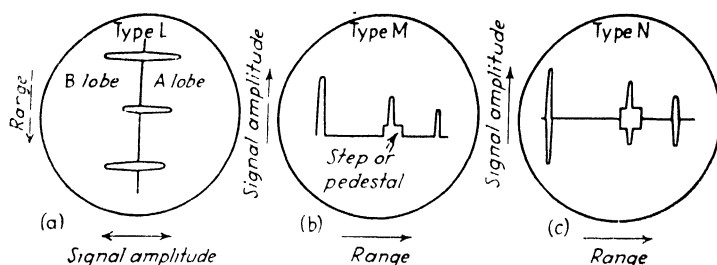



FIG. 27.—Modifications of the type A indicator.

to balance their amplitudes as the array is swung in azimuth or elevation. The traces are horizontal for indicating the signal balance in azimuth or vertical for indicating elevation.

Type L is very similar to type K, except that the two sets of pulses are displayed on opposite sides of the same trace and at the same position so that the range is indicated as in the type A scope, and the azimuth or elevation found by balancing the amplitudes.

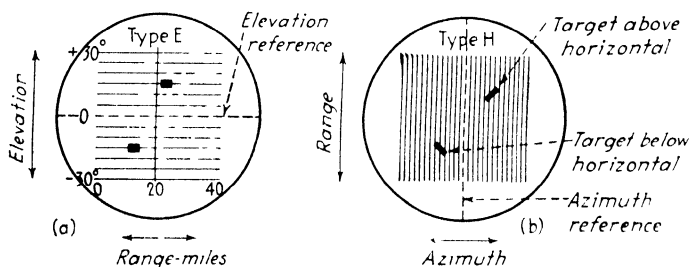



FIG. 28. Modifications of the type B indicator.

A vertical trace is used for azimuth matching, and a horizontal trace is used for elevation matching.

Type M is a conventional type A scope with a raised portion or step that may be positioned under the target signal of particular interest. The coincidence of the pulse and the step is adjusted by a vernier range-finding control, thereby permitting accurate determination of range.

Type N is a combination of types L and M. Type R is an expanded type A indicator, showing a limited segment of range.

20. Modifications of Type B Indicator.—Type E and H indications are modifications of the type B induction, as shown in Fig. 28. In type E the elevation is shown vertically and the range

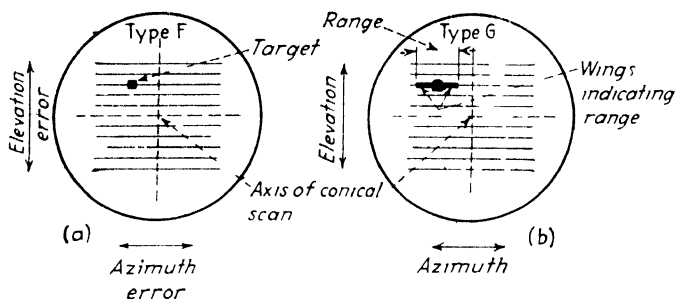



FIG 29 —Modifications of the type C indicator

horizontally. Type H is a conventional type B (azimuth horizontally and range vertically), with the addition of a small line or "semaphore" over the spot indicating the target. The angle made by this extra line is a simple function of the elevation angle. This indicator is used in compact airborne radars where it is essential to provide target information in three dimensions on a single tube. The type B indicator, like the type R indicator, may cover a limited segment of range.

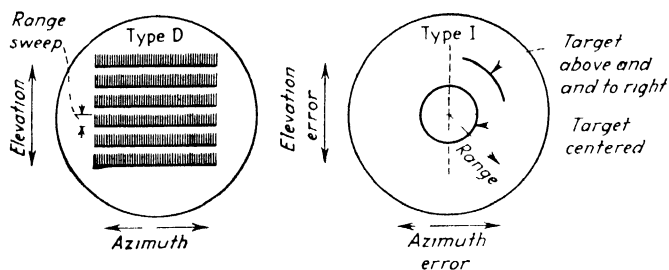



FIG. 30.—Type D, a combination of the type B and type C indicators, and type I, related to the type C

21. Modifications of the Type C Indicator.—Type F and G indications, shown in Fig. 29, are modifications of type C. In type F, the azimuth and elevation indications are error signals, that is, measures of the angular difference between the reference direction (axis of a conical scan) and the target direction. In type G "wings" are added to the type F indication, the length of

the wings being inversely proportional to the range. Type G is another single indicator providing all three target coordinates. The indication closely resembles the appearance of an airplane within view of the radar.

Type D, Fig. 30, is a combination of types B and C. By expanding the horizontal (azimuth) traces, a rough range indication is obtained. This type is not much used. In type I (Fig. 30) the signal is represented by a circular segment whose radius is proportional to range and whose length is proportional to the error of aiming of a conical scan.

22. Summary of Indicator Types.—Table III summarizes the indicators described above.

TABLE III.—RADAR INDICATION METHODS

Type	Echo signal indication	Range indication	Azimuth indication	Elevation indication
A	Vertical	Horizontal	None	None
B	Intensity	Vertical	Horizontal	None
C	Intensity	None	Horizontal	Vertical
D	Intensity	Vertical	Horizontal	Vertical
E	Intensity	Horizontal	None	Vertical
F	Intensity	None	Horizontal (error signal)	Vertical (error signal)
G	Intensity	Wings	Horizontal (error signal)	Vertical (error signal)
H	Intensity	Vertical	Horizontal	Semaphores
I	Intensity	Radial	Arc segment	Arc segment
J	Radial	Circular	None	None
K	Vertical or horizontal	Horizontal or vertical	Pulse match	Pulse match
L	Horizontal or vertical	Vertical or horizontal	Pulse match	Pulse match
M	Horizontal	Step or pedestal	None	None
N	Horizontal	Step or pedestal	Pulse match	Pulse match
P (ppi)	Intensity	Radial	Concentric	None
R (expanded A)	Vertical	Horizontal	None	None

23. The Radar Target.—From the physical point of view the radar target is a discontinuity in the electrical properties of the

medium through which the radio wave travels. The electrical properties are electric conductivity, dielectric constant, and magnetic permeability. If any or all of these quantities change abruptly, that is, within the space of a few wavelengths, a current is induced at the surface of discontinuity when the wave impinges upon it, and the current causes reradiation of the wave. The magnetic permeability of the target is generally disregarded since it is closely the same in all materials except the ferromagnetic substances, and in the latter the conductivity is the predominant characteristic. Hence targets may be classed as conductors, dielectrics, or combinations of both. The induced current is a conduction current (alternating flow of free charges) in the conducting target. It is a displacement current (alternating elastic displacement of bound charges) in the dielectric targets. Otherwise, the two types of target are indistinguishable. The strength of the reradiated wave is generally stronger in the conductors for a given intensity of impinging energy.

The radar target acts as a scatterer of the radio energy falling upon it. Its efficiency as a scatterer is expressed by an effective area of the target variously called the "scattering cross section," "radar cross section," or simply "echo area." This quantity is governed to some extent by the actual area of the target, projected in the direction of the radar, but it is strongly affected by the shape of the target, particularly by the contours of its surfaces and the length and shape of its perimeter, by the frequency of the radar signal, and by the materials of which the target is made. It is a difficult quantity to compute except in the simplest geometrical forms, such as the sphere.

It is customary to treat the echo area of the target as an empirical quantity established by measurement. The measurement, described in more detail in Sec. 109, is performed by measuring the power in watts reradiated by the target in given direction and by dividing this quantity by the power density (watts per square meter) impinging on the target, which may be readily calculated or measured.

From the practical point of view the radar target comprises a category of reflecting objects almost limitless in its scope. Every conceivable object which flies, floats, or "just plain sits" and which may be distinguished electrically from its surroundings constitutes a radar target. In addition to echoes from the radar

target proper, there is the large class of reflections known as "background returns" that arise in the surroundings of the target.

24. Target Types.—In addition to echoes from evident military targets, aircraft or ships, radar echoes may be observed from the

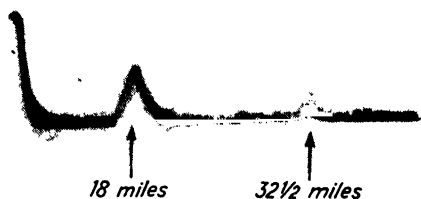



FIG. 31a.—Typical type A indication of particular historical interest. Two Hawker Hart aircraft were observed at the distances shown, on July 24, 1935, the first British observation of aircraft by radar.

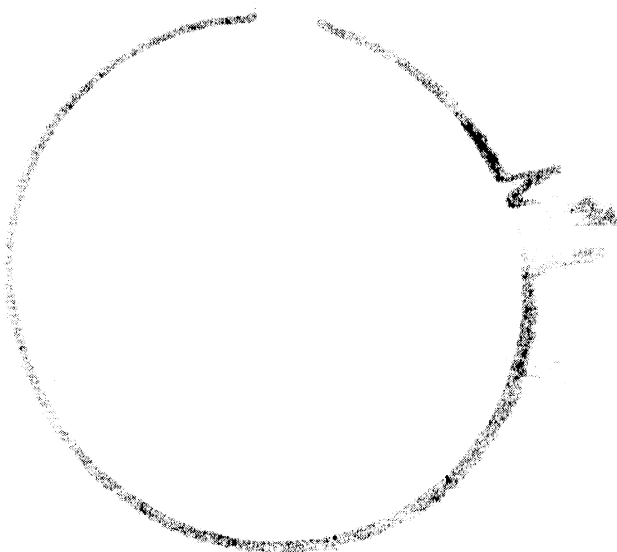



FIG. 31b.—Typical Type J indication, taken from a German Wurzburg antiaircraft gunfire control radar.

splashes or bursts caused by the impact of shells and even from the ionized gases liberated at the muzzle of a gun. Rain, clouds, and other forms of precipitation may be seen, and may partially obstruct the signal, especially when frequencies higher than 3,000 megacycles are employed and when the density of moisture is high.

Reflections from the ionosphere are not observed generally at frequencies above 200 megacycles. However, signals at these frequencies are reflected by boundaries between strata of air of different moisture and heat content. These tropospheric echoes give rise to anomalous propagation effects, discussed in Chap.

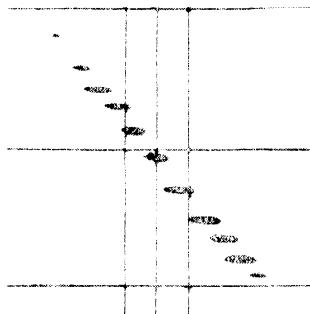



FIG. 31c.—Typical type B (expanded) indication, showing formation of ships separated by 100-yard intervals at a range of 12,900 yards (taken on MPG-1 radar).

IV. Echoes from extensive structures, such as buildings, bridges, cliffs, or mountains are prevalent and may hinder seriously the use of radar. Reflections from such objects are known as "permanent echoes." A source of much mystery in the early days were echoes from an unsuspected source, birds in flight. At the highest frequencies (30,000 megacycles) swarms of insects may be seen. Artillery projectiles may be followed in flight by radars designed for the purpose, and the splash or burst compared by radar with the position of the intended artillery target. Typical targets are shown in Figs. 31 and 32.

Radar targets may be classified as fixed or moving, the fixed target maintaining a fixed position with respect to the radar while the moving target has relative velocity with respect to the radar. A moving target may be distinguished from its back-

FIG. 31*d*.—Typical ppi indication, showing maplike presentation of Nantucket Island from airborne radar, compared with chart of same region (right)

ground, even when no differences in radar reflectance exist, by virtue of the Doppler shift in frequency. The moving target is discussed in detail in Chap. IV, Sec. 111.

25. Distinction between Types of Target.—Identification of the target is of course a primary military requirement. Unfortunately the radar echo does not announce the identity of its

FIG. 32*a*.—Ppi view of Cape Cod taken with 3-cm radar. Concentric circles mark 5-mile intervals. Boston is visible as darkened area, top center. Dark lines are on filter superimposed on screen.

source, but it may give clues based on the average value of the signal strength, the range, and signal strength variations (beats and fades). A small target may be distinguished from a large one by the strength of the echo, provided their ranges are taken into account. Experience with a particular radar will soon indicate the strength of signal to be expected, at various ranges, from targets of different sizes. Variations in signal strength, caused by directive reradiation or by wave interference, are characterized by

the rapidity and depth of the fading. Such signal variations may be studied on the Type A indicator. Ships exhibit deeply fading echoes of slow period, whereas the echo from a fast-flying aircraft may vary so rapidly that the variation is evident merely as a shimmering of the top of the pulse (as viewed on the A scope). A highly characteristic beating effect is produced by "window," a cluster of half-wavelength tinfoil conductors dropped from aircraft to confuse enemy radar operators and hide the activities of aircraft that may follow.

26. Background Returns.—Whenever a radar beam is directed at the surface of the earth, land, or water, a multiplicity of reflec-

FIG. 32b.—Mediterranean coastline as revealed in a series of ppi views. Approximately ten presentations are shown, assembled in proper relative positions.

tions is produced that, coming from a large area, is generally strong compared with the reflection from the target itself. One evident limitation produced by ground and sea return is the limitation of minimum range in airborne radars. Reflections from the earth beneath the aircraft are generally very strong compared with those from targets viewed in the horizontal direction, so much so that ground return is generally present even when highly directive aerials are used. The presence of the dense ground echo, at a range corresponding to the aircraft's altitude, prevents observing targets horizontally at the same range. The ground return extends with decreasing intensity outward to the horizon as viewed from the aircraft, although

targets may often be seen through the ground return arising from points outside the nadir of the aircraft position

27. Terrain Echoes.—An important application of sea and ground return is the distinction of terrain features, such as coastlines, rivers, mountain ranges, and cities, by variations in radar reflectance (see Figs 32 and 33) Such distinctions are

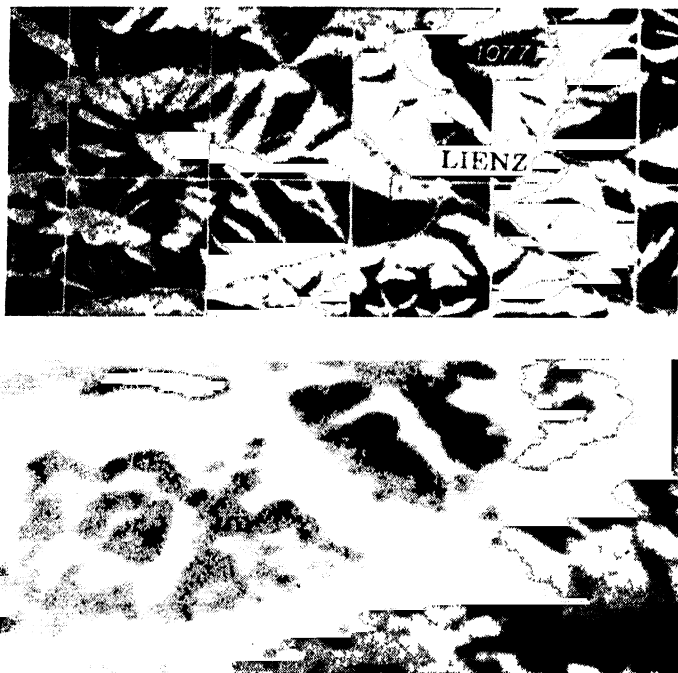



FIG 33*a* —Relief map (top) of mountainous terrain, and ppi radar view of same area (bottom) The radar illumination is from the top of the picture

based principally on the roughness vs smoothness of the terrain, rather than on inherent electrical characteristics Thus when viewed obliquely a shore line scatters more energy than the sea, despite the fact that the sea is a better conductor.

Such terrain echoes have been used with outstanding success for navigation, both marine and aerial, and for sighting on bombing objectives. Success in the latter case depends upon the use of a very high frequency and extremely sharp beam to resolve the reflectance differences between the bombing objective and its surroundings A 30,000-megacycle radar of this type is capable

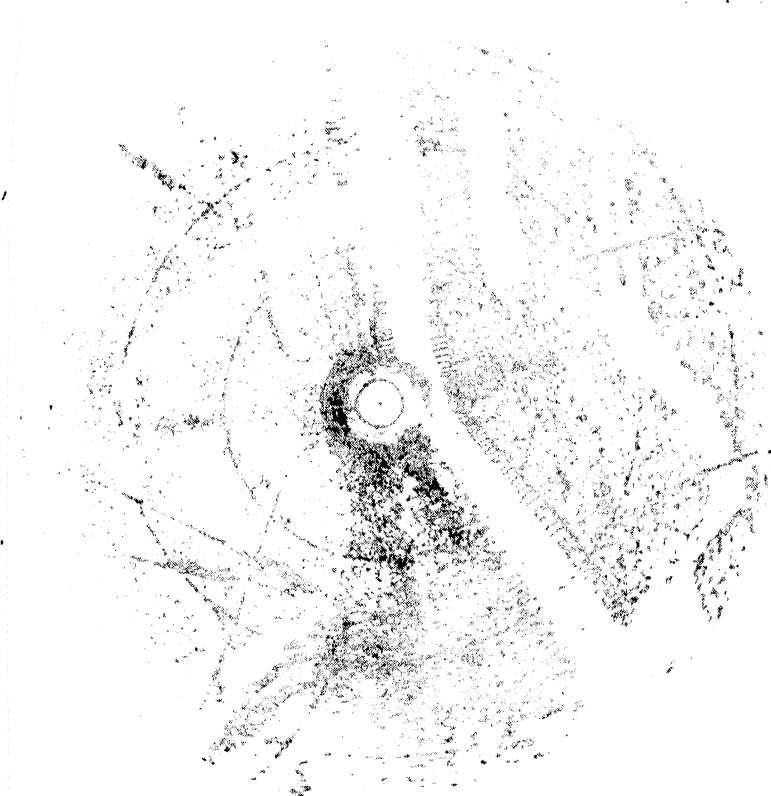



FIG. 33b.—Ppi view of Manhattan Island and environs, taken with 1-cm radar at a height of 10,000 feet. Note individual docks along Hudson River (center) Central Park, and railroads in Jersey meadows (left).

of resolving separately the docks along a waterfront from an aircraft flying at 10,000 ft (Fig. 33b).

28. The Free-space Radar Equation.—The foregoing description of radar methods serves to indicate the general approach. But before further conclusions can be drawn regarding radar design and performance, it is necessary to establish the quantitative relations between transmitted power and receiver sensitivity, the target characteristics and position relative to the radar. The equation that links these factors may be termed appropriately the "radar equation" since its formulation is unique to the radar system. There are two basic forms of radar equation, depending on the type of radio propagation present;

and many minor variations of each have been derived. The simplest and most fundamental is the *free-space radar equation*, which governs the radar signal when it is propagated between a radar and a target in otherwise empty space. The *surface-propagation radar equation*, which takes into account the reflection of the waves from the earth's surface between radar and target is derived in Chap. IV, in connection with a detailed statement of radar propagation effects.

FIG. 34.—Geometry of free-space radar propagation from point-source radiator, during outward passage of radar signal.

The free-space radar equation relates the power radiated from the radar (transmitter output and antenna gain) to the power delivered to the terminals of the receiver, taking account of the dispersion in space and of the reflecting properties of the target. It is derived as follows:

Consider a transmitter that delivers a peak power of P_t watts to the radiator. If the radiator is an isotropic source, this power is radiated equally in all directions, filling a sphere of constantly increasing radius (Fig. 34). The density of the power at a distance r meters is equal to the power radiated divided by the area $4\pi r^2$ of the sphere concentric on the radiator. In practice the

FIG. 35.—Same as Fig. 34, except radiator is directive.

radiator is fitted with a reflector that increases the power density in particular directions at the expense of the power density in the other directions. The directivity is described by the power gain of the radiating system. The gain of the antenna is not uniform in all directions but displays a maximum generally along the axis or meridian plane of the beam, as shown in Fig. 35. The maximum power gain of the radiator is denoted by G_0. The

power density S_t radiated in the direction of maximum gain is, then

$$S_t = \frac{G_0 P_t}{4\pi r^2} \qquad \text{watts per sq m} \tag{7}$$

When the wave encounters the target, at range r, the target absorbs and reradiates an amount of power equal to the impinging power density S_t times the echo area of the target. This echo area has been mentioned and the difficulties of computing it have been suggested. Here we merely assign its value as σ sq m, which we may assume has been determined empirically.

FIG. 36.—Geometry of free-space radar propagation, during return passage of signal from target to radar.

The target is, then, a source of radio waves of power σS_t. It is evident that the power reradiated is not uniform in all directions, but the value σ is taken as that effective in the direction toward the radar.

The wave, in traveling the distance r back to the radar, spreads through a sphere of radius r (Fig. 36). Hence it displays, at the radar, a power density of

$$S_r = \frac{\sigma S_t}{4\pi r^2} \qquad \text{watts per sq m} \tag{8}$$

Substituting Eq. (7), this is

$$S_r = \frac{G_0 P_t \sigma}{(4\pi)^2 r^4} \qquad \text{watts per sq m} \tag{9}$$

This power density is gathered by the effective absorption cross section of the receiving antenna, A_0 sq m, and generates a power at the receiver input of

$$P_r = A_0 S_r \qquad \text{watts} \tag{10}$$

Substituting Eq. (9), this becomes

$$P_r = \frac{A_0 G_0 P_t \sigma}{(4\pi)^2 r^4} \qquad \text{watts} \tag{11}$$

This is the general form of the free-space radar equation. It gives the received power level in terms of the transmitted power,

the properties and distance of the target, and the properties of the transmitting and receiving antennas. Equation (11) may be recast to express the range of the target in terms of the other quantities.

$$r = \sqrt[4]{\frac{P_t A_0 G_0 \sigma}{P_r (4\pi)^2}} \quad \text{meters} \tag{12}$$

The form of the equation of most practical interest is that expressing the maximum range at which detection of a given target is possible. For a given transmitter power P_t, antenna system $A_0 G_0$, and target echo area σ, the range is limited by the minimum received power $P_{r\min}$ that will generate a barely discernible response in the radar indicator. The maximum range is given by

$$r_{\max} = \sqrt[4]{\frac{P_t A_0 G_0 \sigma}{P_{r\min} (4\pi)^2}} \quad \text{meters} \tag{13}$$

A surprisingly large amount of information is given by this equation. It states that the maximum range of a radar depends on the fourth root of (1) the transmitter power, (2) the antenna gain and absorption area $G_0 A_0$, (3) the target echo area σ, and (4) the inverse $1/P_{r\min}$ of the minimum discernible received power ($1/P_{r\min}$ is the receiver power sensitivity).

Equation (13) also reveals that equal effects on the maximum range may be achieved by increasing the transmitter power P_t or the power sensitivity $1/P_{r\min}$ of the receiver. It indicates the apparent facts that the range is greater when the antenna gain is high and when the target has a large effective area.

The fourth root modifies these relationships in a very important way. It indicates that a given increase in maximum range is obtained only at great expense in additional transmitter power and antenna gain and that the range possible when viewing large targets is only slightly greater than that possible with small targets. Doubling the value transmitted power, for example, increases the maximum range by $\sqrt[4]{2} = 1.19$ times. Doubling the value of A_0, G_0, or σ (or halving the value of $P_{r\min}$) has the same small effect, a 19 per cent increase in maximum range In engineering terms, a 12 db increase in transmitted power or receiver sensitivity is required to double the maximum range.

29. Influence of Operating Wavelength.—The radar equation expressed in Eq. (13) gives no clue as to the dependence of radar

performance on the operating wavelength or frequency. This information is implicit in the quantities A_0, G_0, and σ. The radiator gain G_0 and absorption cross section A_0 are readily computed, to a fair approximation, from the operating wavelength and the area of the reflector surface.

The gain G_0 is defined as the ratio of the solid angle (4π steradians) subtended by a point-source (isotropic) radiator to the solid angle subtended by the beam. The angle subtended by the beam depends evidently on the form of the radiator. One of the commonest forms, discussed fully in Chap. IV, is the circular paraboloidal reflector, whose diameter is D m and area across the face is $A = \pi D^2/4$ sq m. The beam produced is of circular cross section and occupies a beam width, between half-power points, of

$$b = \frac{\lambda}{D} = \frac{\lambda}{2\sqrt{A/\pi}} \qquad \text{radians} \tag{14}$$

where λ is the operating wavelength in meters.[1] The maximum gain of such a beam radiator, relative to an isotropic radiator, is

$$G = \frac{8\pi A}{3\lambda^2} \tag{15}$$

The radiator proper is customarily a dipole mounted at right angles to the direction of maximum gain. The dipole has an inherent power gain, in that direction, of 3/2 over an isotropic radiator. Hence the total maximum gain G_0 of the radiator is

$$G_0 = \frac{3}{2} \times \frac{8\pi A}{3\lambda^2} = \frac{4\pi A}{\lambda^2} \tag{16}$$

The gain of a radiator may also be expressed as

$$G_0 = \frac{KA}{\lambda^2} \tag{17}$$

where K is an empirical constant that takes account of losses not contemplated in the theoretical derivation. The measured value of K is 6 to 8 for paraboloidal reflectors and 10 to 12 for mattress-type dipole arrays.

[1] This is the minimum possible beam width. In practice the beam width of a circular parabolic radiator is about 25 per cent greater than the minimum value. See Sec. 97, Chap. IV.

The absorption cross section A_0 is related to the maximum receiver antenna gain G_r by the following

$$A_0 = \frac{G_r\lambda^2}{4\pi} \qquad \text{sq m} \tag{18}$$

Since virtually all radars use a single reflecting system and dipole, or dipoles, for transmitting and receiving, $G_r = G_0$, and

$$A_0 = \frac{G_0\lambda^2}{4\pi} \qquad \text{sq m} \tag{19}$$

Substituting Eq. (16) for G_0

$$A_0 = A \qquad \text{sq m} \tag{20}$$

which is simply the area of the reflector. Also

$$A_0 = \frac{KA}{4\pi} \qquad \text{sq m} \tag{21}$$

which takes account of the empirical efficiency constant K.

The dependence of σ on the operating wavelength cannot be generally stated except for targets of simple shape. Accordingly the quantity σ is retained without further analysis, with the understanding that its contribution to the radar equation does in fact vary with the operating frequency, but in a manner which is so dependent upon the particular target and its aspect that no general conclusions can be drawn.

Returning now to the radar equation for maximum range (13) and substituting Eqs. (16) and (20) for G_0 and A_0, we obtain

$$r_{\max} = \sqrt[4]{\frac{P_t A^2 \sigma}{P_{r\min}\lambda^2(4\pi)}} \qquad \text{meters} \tag{22}$$

This equation indicates that the maximum range increases inversely as the square root of the wavelength, or directly as the square root of the frequency, provided that we can choose the area A of the reflector without regard to frequency. Thus if we can select a convenient value of A and let it remain fixed, the maximum range will increase slowly, all other factors remaining unchanged, as the frequency is increased. Unfortunately, all other factors do not remain fixed as the frequency is increased into the thousands of megacycles. The available transmitter

power and the receiver power sensitivity both decrease, and the range may thereby decrease, at the higher frequencies.

One important aspect of radar design is the beam width λ/D. As the wavelength is decreased (frequency increased) with a given area of reflector, the beam becomes narrower. A practical limit is reached, with a fixed value of A, when the beam becomes so narrow that it is impractical to scan a large region of space in a reasonable time. Hence it is usual to consider the effect of frequency on maximum radar range, when the beam width is fixed. To show this Eq. (22) is rewritten, substituting Eq. (14), as

$$r_{max} = \sqrt[4]{\frac{P_t \pi \sigma \lambda^2}{P_{rmin} 64 b^4}} \quad \text{meters} \tag{23}$$

Thus, when the beam width is fixed, the maximum range *increases* with the square root of the operating wavelength. It is necessary, of course, to use large reflectors to produce a suitably narrow beam at low frequencies, but when this is done the range achieved is greater than in comparable h-f radars of the same beam width. The greater available transmitter power and higher receiver sensitivity contribute to the longer ranges achieved at lower frequencies.

30. Influence of Receiver Characteristics.—The quantity P_{rmin} that measures the power sensitivity of the receiver must be examined in detail to relate the maximum radar range to receiver characteristics. P_{rmin} is the minimum reflected signal power to which the receiver will give a discernible response against noise. The quantity depends, therefore, on (1) the amount of noise visible in the indicator, and (2) the amount of signal necessary to give an indication just distinguishable against the noise. The first factor, the noise present, arises from many sources, the most important of which is circuit (thermal) noise generated in the receiver. The second factor, the minimum signal required to produce a response against the noise, is fundamentally a psychological quantity, since the act of discerning the signal involves the eye and mind of the observer. Before the least discernible signal can be specified in a particular instance, it is necessary to know the type of indicator employed (for example, type A, B, C, or ppi), the pulse rate, the persistence and contrast characteristics of the phosphor, and similar factors involved in

transferring the signal voltage into a visual indication. The subject of signal-to-noise ratio in a radar receiver is thus a very complicated matter. A more complete account of the theory and practice of noise measurement is given in Chap. II, Sec. 60.

The quantity P_{rmin} is measured by a standardized procedure under the condition that the rms value of the signal output shall just equal the rms value of the noise actually present in the receiver output (signal-to-noise power ratio of 1:1). Such a signal-to-noise ratio represents a discernible response on an adequately designed type A indicator (Fig. 37). The power so measured is compared to the irreducible minimum of noise associated with the system, that is, the noise arising externally to the receiver. The ratio of the signal (necessary to equal all noise present) to the irreducible noise is called the "noise figure" of the receiver in question. The higher the noise figure, the poorer the performance of the receiver. Thus an ideal receiver (one that has no source of noise within itself) when completely overcoupled to the antenna has a noise figure of 1, since the signal is then competing only with the external noise. An ideal receiver whose input impedance matches the radiation resistance of the antenna has a noise figure of 2, since additional noise is added by the input impedance. Practical receivers have noise figures from about 3 to 100, depending on the magnitude of other sources of noise (particularly the converter stage) within the receiver.

FIG. 37.—Type A indication showing noise on base line. The smallest signal (at right) has a signal to noise ratio of approximately 1:1, representing close to the minimum detectible level.

In consequence of the above definitions we write

$$P_{rmin} = nkT\,\Delta f \qquad \text{watts} \tag{24}$$

where n is the receiver noise figure and $kT\,\Delta f$ is the irreducible available noise power, k is Boltzmann's constant, 1.374×10^{-23} watt per deg per cycle, T is the absolute temperature in degrees Kelvin (degrees centigrade plus 273°), and Δf is the effective bandwidth of the system. As shown in detail in Sec. 60, $kT\,\Delta f$, the "available" external noise power, is one-quarter of the noise

power ($4kT\,\Delta f$) generated by the random motion of charges in space and developed across the radiation resistance of the antenna. If the radar were in a perfect vacuum, no such charges would be present and the external noise of the receiver would be zero. In practice the thermal energy of molecules in the atmosphere produces the noise of space. The equivalent temperature of space T is commonly taken as room temperature (290°K) but recent measurements have proved it to be much lower than this. Since T of space is not definitely known, the limit of receiver sensitivity is not, as yet, a definitely known quantity. For purposes of standardization T is taken as 290°K, and $kT\,\Delta f$ equals 4.1×10^{-21} watt per cycle of effective band width. Thus a radar receiver using a 1-megacycle bandwidth must be sensitive to a power of about 10^{-13} watt, or $\frac{1}{10}$ $\mu\mu$w.

No mention has been made of man-made interference or natural atmospherics in the external noise discussed above. Such noise is usually the limiting factor in vhf (very high frequency) and lower frequency communication. In the uhf (ultrahigh frequencies) and shf (superhigh frequencies) regions employed in radar, however, these components of external noise are considerably less than the thermal noise generated across the radiation resistance and hence may be lumped together in the quantity $kT\,\Delta f$.

The fact that the thermal noise power present in radar reception increases in direct proportion to the effective bandwidth has important consequences. In particular it indicates that the important quantities affecting maximum range are the *average power* (rather than the peak power) and *pulse rate* of the transmitter, or alternatively the *energy* in each pulse (peak power times pulse width). To show this, we substitute $nkT\,\Delta f$ for $P_{r\,min}$ in Eq. (22),

$$r_{max} = \sqrt[4]{\frac{P_t A^2 \sigma}{nkT\,\Delta f\,4\pi\lambda^2}} \quad \text{meters} \tag{25}$$

This equation states that the maximum range decreases as the bandwidth is increased. As is shown in detail in Chap. II, the bandwidth required varies inversely as the pulse width. In particular, for optimum peak signal-to-noise ratio, the bandwidth between half-power points (3 db down) should be approximately $1/d$ megacycles where d is the pulse width in microseconds.

Substituting $\Delta f = 1/d$ in Eq. (25) we obtain

$$r_{max} = \sqrt[4]{\frac{P_t d A^2 \sigma}{nkT4\pi\lambda^2}} \quad \text{meters} \tag{26}$$

Hence the fourth root of the quantity $P_t d$ (the peak power times the pulse width or the energy of the pulse) determines the range of the system, rather than the peak power alone. The average transmitted power P_{av} is related to the pulse energy by the pulse rate. According to Sec. 6, page 21,

$$P_{av} = \frac{P_t d}{i} = P_t d f_p \quad \text{watts} \tag{27}$$

and

$$\frac{P_{av}}{f_p} = P_t\, d \quad \text{watts} \tag{28}$$

It follows that the energy in each pulse may be expressed as the quotient of the average power by the pulse rate. Neither the peak power nor the average power, by itself, determines the maximum range on a given target with a receiver of given noise figure. Information on the pulse width or pulse rate, respectively, must also be available.

The above analysis is made plausible by the following reasoning: The range of a radar may be increased by increasing the peak power of the pulse, provided that the pulse width is not proportionately decreased in the process. If the pulse width is decreased in proportion, the necessary bandwidth and the thermal noise power increase in proportion, and thus no benefit is derived from the increase in peak power.

If we increase the peak power without changing the pulse width, we must either increase the average power or decrease the number of pulses per second proportionately. For the reasons discussed in Sec. 6, it may not be possible to reduce the pulse rate, because the pulse rate must be high enough, in relation to the scanning rate, to deliver pulses to all points in the scanned space. Hence, the pulse width and pulse rate remaining fixed, the range of a radar system is increased only by increasing the average power transmitted. This is the reason for the statement (page 9) that the average power is a better criterion of radar performance than peak power. In other words, high peak power is not used for its own sake. It is used, rather, to obtain sufficient

average power for a given maximum range, when the pulse specifications have been set by other considerations (maximum range, minimum range, required resolution in range, allowable scanning motion per pulse, and allowable scanning interval).

31. Numerical Examples Based on the Free-space Radar Equation.—Figures 38 and 39 are based on typical values of the

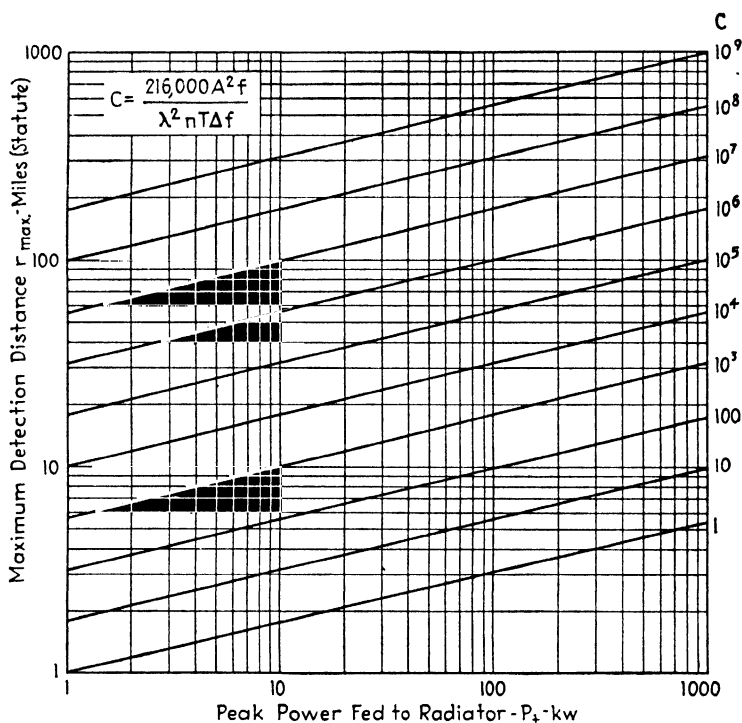



FIG. 38.—Relationship between transmitter power and maximum range for various values of the system constant C, based on free-space radar equation.

quantities in Eq. (25). In the interest of generalization, the equation is written in two steps, and a plot made of each.

$$r_{max} = \sqrt[4]{CP_t} \qquad \text{statute miles} \tag{29}$$

and

$$C = \frac{MNA^2\sigma}{nkT\,\Delta f\,4\pi\lambda^2} = \frac{MNA^2\sigma\lambda^2}{nkT\,\Delta f\,4\pi\lambda^4} \qquad \text{miles}^4 \text{ per watt} \tag{30}$$

where all the quantities have been defined except M, a constant to convert r_{max} from meters to statute miles, and N, a factor to

provide a margin of safety against losses (not previously considered) in the radiator and transmission line system, particularly in the t-r switch. The quantity C may be thought of as a system constant describing the radar and target in all aspects except the peak power.

Equation (29) is plotted in Figure 38, for various values of C from 1 to 10^9, for peak power P_t from 1 kw to 1,000 kw, and for ranges r_{max} from 1 to 1,000 miles.[1] Figure 38 shows at once that the peak power has little effect on the maximum range for a given system constant C. The slope of the curves is one-fourth (fourth root on log-log coordinates). Inspection shows that 16 times (12 db) additional power is required to double the range and 10,000 times (40 db) additional power is required to increase the range by a factor of 10, for example, from 10 to 100 miles. Figure 38 shows the similarly small effect on maximum range of the system constant, which must also increase 10,000 times to cause the range to increase 10 times.

The system constant C (Eq. 30) is plotted in Fig. 39 with appropriate values of the constants for five frequencies used in radar equipment: 200, 600, 3,000, 10,000 and 30,000 megacycles. The corresponding values of λ are 150, 50, 10, 3, and 1 cm. The figure shows the typical values of n and Δf in each case. The other constants are $M = (1{,}609)^4$, $N = 0.25$, $k = 1.37 \times 10^{-23}$, $T = 300$, and $\sigma = 10$ sq m. The ratio A/λ^2, the area of the reflector in square wavelengths, is taken from 1 to 1,000, and typical values of A used at each frequency are identified as dots on the lines. The target echo area is taken as 10 sq m since this is typical value (a fighter aircraft). Other values of σ may be taken into account by multiplying C by $\sigma/10$, where σ is the target echo area in square meters.

The use of the plots is evident. For example, consider a 10-cm (3,000-megacycle) radar, whose radiator has a reflector area of 100 square wavelengths (1 sq m, 45-in. paraboloid). Then $A/\lambda^2 = 100$, the beam width is 5.7 deg, and $C = 11{,}000$,

[1] It should be noted that free space transmission, on which this analysis is based, seldom occurs at distances beyond 150 miles (the distance to the horizon from a plane flying at 20,000 ft). The free-space maximum range is equal to one-half the maximum range, attainable under ideal conditions, when surface propagation is present, so the curves have general significance (see Chap. IV, Sec. 102).

for the assumed values of the constants given above. A peak power of 50 kw will produce a maximum range of 28 statute miles on a target of 10-sq m echo area.

It must be remembered that this is the maximum free-space range, using a type A indicator of optimum design, when the radar beam points in a fixed direction (that is, scanning losses

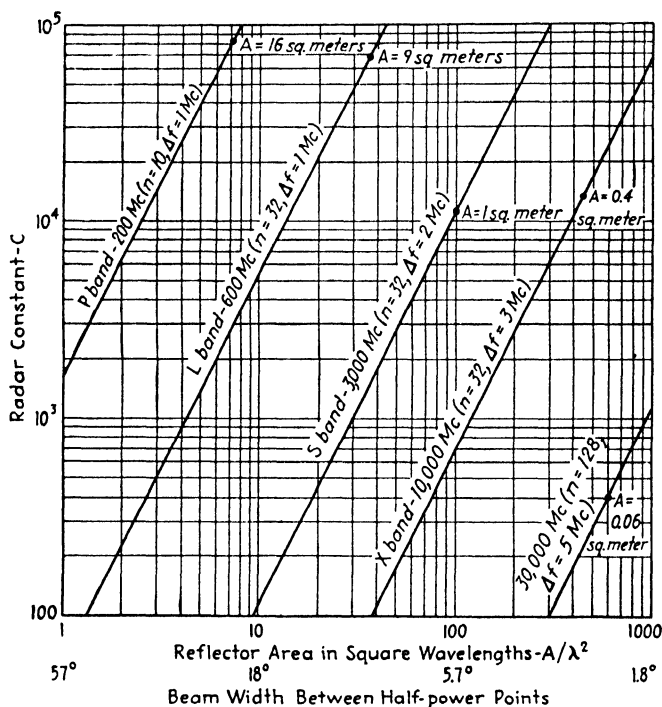



FIG. 39.—The radar system constant in terms of radiator size and other constants typical of operation at various frequencies from 200 to 30,000 mc.

are not considered). Experience with typical equipment shows that the predicted ranges are observed in fact.

In practice, range measurements are taken as measures of the target echo area σ, since σ cannot be measured conveniently in any other way. The consistency of results with a given target, viewed at different ranges, supports the analysis given above.

32. Summary.—The radar equation shows that the significant items in a radar system are as follows:

1. Transmitter
   a. Pulse energy ($P_t d$ or P_{av}/f_p)
   b. The carrier frequency or wavelength (λ)
   c. Antenna gain ($G_0 = KA/\lambda^2$)
2. Receiver
   a. Noise figure (n)
   b. Bandwidth (Δf)
   c. Antenna gain ($G_r = G_0$)
3. Target
   a. Echo area (σ)
   b. Range (r)
   c. Angular deviation from axis of beam (ϕ and θ)

This summary suggests that there are two main divisions of radar engineering: the pulse aspects and the r-f aspects. The pulse aspects cover pulse generation according to prescribed specifications, pulse transmission, bandwidth and noise, pulse amplification at intermediate and video frequencies, and indication of the pulses. These matters are considered from the fundamental point of view in Chap. II. The r-f aspects include the generation and modulation of r-f energy, radiators and their characteristics, transmission in lines and guides, propagation, the echo area of the target, scanning, and reception of r-f energy. Chaps. III and IV consider the r-f technology from a fundamental point of view. Part II of this book discusses practical embodiments of these principles in component parts and complete systems.

Bibliography

Joint Board on Scientific Information Policy: "Radar, A Report on Science at War," Government Printing Office, Washington, D.C., August, 1945.

SMITH-ROSE, R. L.: Radiolocation, Principles and Origins, *Wireless World*, **51** (2), (February, 1945); **51** (3), 66 (March, 1945).

FINK, D. G.: The Radar Equation, *Electronics*, **18** (4), 92 (April, 1945).

———: Radar Specifications, *Electronics*, **18** (11), 116 (November, 1945).

APPLETON, E.: Scientific Principles of Radiolocation, *Jour. Inst. Elec. Eng.* (*London*), **92**, Part 1, 340 (September, 1945).

COLTON, R. B.: Radar in the United States Army, *Proc. I.R.E.*, **33**, **740** (November, 1945).

DUBRIDGE, L. A.: History and Activities of the Radiation Laboratory of the Massachusetts Institute of Technology, *Rev. Sci. Inst.*, **17**, 1 (January 1946).

Evolution of Radar, *Engineering* (*London*), **160**, 154, 176, 183, 205, 236, 244, 265, 285 (Aug. 24 to Oct. 12, 1945).

MOFENSON, J.: Radar Echoes from the Moon, *Electronics*, **19** (4), 92 (April, 1946).

RIDENOUR, L. N.: Radar in War and Peace, *Elec. Engineering*, **65**, 202–207 (May, 1946).

DEWITT, J. H., JR.: Technical and Tactical Features of Radar, *J. Franklin Inst.*, **241**, 97–123 (February, 1946).

SCHNEIDER, E. G.: Radar, *Proc. I.R.E.*, **34**, 528–578 (August, 1946).

BERKNER, L. V.: Naval Airborne Radar, *Proc. I.R.E.*, **34**, 671–706 (September, 1946).

WATSON-WATT, SIR ROBERT: Evolution of Radiolocation, *Engineer*, **161**, 319, 330 (Apr. 5, Apr. 12, 1946).

HAEFF, A. V.: Minimum Detectable Radar Signal and Its Dependence upon Parameters of Radar Systems, *Proc. I.R.E.* **34**, 857 (November, 1946).

CHAPTER II

PRINCIPLES OF PULSE GENERATION AND TRANSMISSION

33. Pulse Methods in Radar.—The study of radar methods and equipment must be based on a thorough understanding of pulse techniques. These techniques fall into five categories: (1) pulse generation, (2) modification of shape and interval, (3) transmission at video frequencies, (4) modulation and demodulation of r-f (radio-frequency) sources, and (5) transmission at carrier frequency or intermediate frequency.

It is evident that before a pulse can be used it must be generated in some fashion. But it is rare that a pulse generator can produce directly pulses that have all the characteristics required. Rather, it is generally necessary to modify the pulse width or the pulse shape. It may also be desirable to change the pulse rate or otherwise modify the size and regularity of the intervals between pulses.

When a series of pulses of the desired characteristics has been generated, it is necessary to transmit the pulse at video frequencies, before modulating the r-f source. Such transmission generally involves a number of video amplifiers, coupling networks, input and output terminations, etc., all of which must be designed to transmit the pulses without undue impairment of their shape and to deliver them at a voltage or power level suitable to the task to be performed.

The pulses thus delivered are applied to an r-f source to modulate it. At this point the pulse technique, strictly speaking, is replaced by the r-f technique and the succeeding processes of propagation and reception. However, the principles of transmitting pulses at carrier and intermediate frequencies are so markedly similar to those governing video transmission that they are conveniently discussed together.

The pulses reappear, as such, after demodulation in the receiver. At this point a new factor assumes importance: the signal-to-noise ratio. Throughout the system, compromises

must be adopted to minimize the noise without undue impairment of the pulse shape and amplitude.

The foregoing is the succession of events relating to the radar signal itself. Pulses are used, in addition to the signal function, for many purposes in radar equipment. They coordinate the timing of the system, initiate sweep circuits, act as range markers and other calibration signals, and control the gain of the receiver. Pulses used for calibration and control are similar to the pulse signal, except that they occur generally at video frequencies, rather than at carrier or intermediate frequencies.

In the following sections the principles underlying the design of pulse systems are discussed. The factors underlying pulse generation and modification are generally the simplest and most familiar, since they comprise the transient behavior of individual circuit elements and vacuum tubes. Accordingly the transient response of simple circuit elements is taken as the starting point.

34. Transient Response of Circuit Elements.—The passage of a pulse through a transmission system is observed in terms of the *responses* of the circuit elements and vacuum tubes that comprise the system. There are two closely related circuit responses: (1) the current that flows in a circuit when a voltage is applied across it and (2) the voltage that appears across a circuit when a current is forced through it.

The relationship between current (I) and voltage (E) is expressed as a ratio called "impedance" ($Z = E/I$), or "admittance" ($Y = I/E$). The impedance of a circuit is a familiar concept. In communications practice it is usual to express the impedance as a function of frequency. This is done by measuring or computing the ratio between applied voltage and resulting current at all frequencies of interest, for example, throughout the v-f (video-frequency) range.

When a voltage waveform is applied to the circuit, the response of the circuit is determined by analyzing the applied waveform into its sinusoidal components and by treating the problem as though each voltage component were applied separately. The current responses to the several voltage components are computed or measured. The sum of the responses constitutes the current waveform that flows in response to the applied voltage. This is the "steady-state" method of measuring or computing

circuit responses in linear systems. The same method is used to obtain the voltage response to an applied current.

The steady-state approach to circuits is convenient when the applied voltages are sinusoidal, as they are in power practice and in certain a-f (audio-frequency) problems. The method is applicable to any applied waveform and it may be used when the applied voltage or current is a pulse. But when pulses are applied to simple circuits, consisting of one or two circuit elements, the steady-state method may prove cumbersome. In this case it is simpler, and considerably more informative, to measure or compute the *transient response* of the circuit.

The transient response of a circuit is its response to a suddenly applied voltage or current. The suddenly applied voltage is zero up to the time of application. At this instant, the voltage instantaneously assumes a finite value, which is maintained across the circuit for an indefinite period thereafter. The current that flows in response to the suddenly applied voltage in general does not have the same waveform as the applied voltage. The shape of the current waveform can be computed in many simple cases by direct integration of the differential equations that describe the circuit elements, as we shall show in subsequent paragraphs.

Knowledge of the transient responses of simple circuit elements and combinations is of great value in explaining or predicting the reaction of a circuit to a pulse. The simplest form of pulse, the rectangular pulse shown in Fig. 40, is in fact the sum of two transients, a sudden increase at the beginning of the pulse followed by an equal and opposite sudden decrease at the end of the pulse. The response of a circuit to a rectangular pulse is thus the sum of two transient responses, one positive, the other negative, separated by an interval equal to the width of the pulse.

It is evident that the transient analysis comes much closer to describing the physical reality of the circuit action, when pulses are applied, than does the steady-state analysis. But when complicated circuits are considered, it may not be possible to integrate the circuit equations simply. In this case, it may prove more convenient to resort to the steady-state formulation.

In what follows, the voltage and current sources are assumed to possess perfect regulation. When voltage is applied across a circuit, the voltage source is assumed to have zero internal

impedance, so that the voltage applied does not change no matter what current may flow through the voltage source. This condition is approximated, for example, by a low-impedance vacuum tube (for example, cathode coupled) feeding a high impedance circuit. Likewise, when current is applied to a circuit the current source is considered to have infinite internal impedance, such that the current does not change no matter what value of voltage may appear across the current source. This condition is approximated by a high-impedance vacuum tube (for example, pentode) feeding a low impedance. These assumptions, by divorcing the

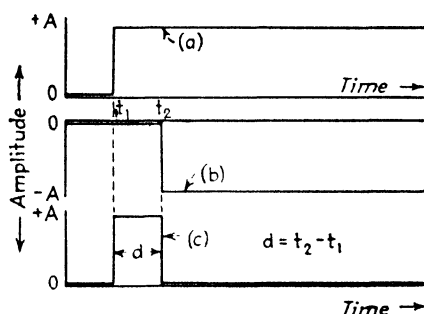



FIG. 40.—Development of basic rectangular pulse from positive and negative unit steps.

internal characteristics of the source, confine our attention to the response of the circuit elements. In many practical cases the internal resistance of the generator may be added to the resistance of the connected circuit.

35. Transient Response of Single Circuit Elements. Resistance.—A pure resistance R ohms has negligibly small storage properties, that is, it does not possess appreciable inductive or capacitive reactance. Accordingly, when a sudden change in voltage or current is applied to a pure resistance, it responds instantaneously, without the delay incident to energy storage. When a sudden change in voltage is applied, the magnitude of the current change is related to the magnitude of the applied voltage change by Ohm's law

$$i = \frac{e}{R} \quad \text{amp} \tag{31}$$

where e is the magnitude of the applied voltage change in volts

and R is the value of the resistance in ohms. Figure 41 shows the response graphically. Essentially, then, the response of a pure resistance to a voltage pulse (two successive and opposite transients) is a current pulse of the same shape and of amplitude governed by Eq. (31).

When a sudden change in current is forced through a resistance, a corresponding sudden change in voltage appears across the resistance. The voltage response has the same shape as the current applied. Thus current and voltage may be interchanged as cause and effect without changing the waveforms. The amplitudes of current and voltage are governed in either case by Eq. (31).

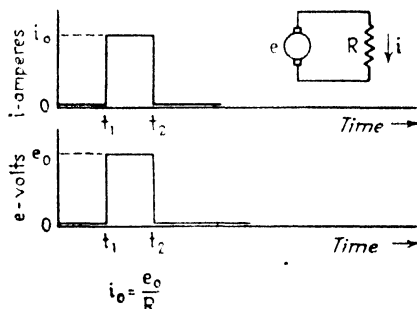



FIG. 41.—Current and voltage responses in a resistive circuit.

Inductance.—A pure inductance stores all the electric energy fed to it in the magnetic field surrounding the conductor. The magnetic field cannot be established instantaneously. The back emf, induced by the magnetic field, opposes the applied voltage and is proportional to the time rate of change of the current in the inductance. Since the voltage applied during the pulse is constant, the current increases continuously and linearly during the pulse.

The differential equation of an inductance is

$$e = L\frac{di}{dt} \qquad \text{volts} \tag{32}$$

where di is the differential change in current occurring during the corresponding differential time dt and e is the applied voltage. When this equation is integrated for a constant value of e it

becomes[1]

$$i = \frac{e}{L} t \quad \text{amp} \tag{33}$$

This equation states that the current response of a pure inductance L henries to e volts increases linearly with time at the rate e/L amp per sec. When the sudden change in voltage is applied at time $t = 0$, no current flows, but it begins immediately thereafter and increases so long as the voltage is applied. When the pulse is terminated, the current begins to decrease linearly, with slope e/L, until the current is zero. Thus the response of a pure inductance to a rectangular voltage pulse is a triangular

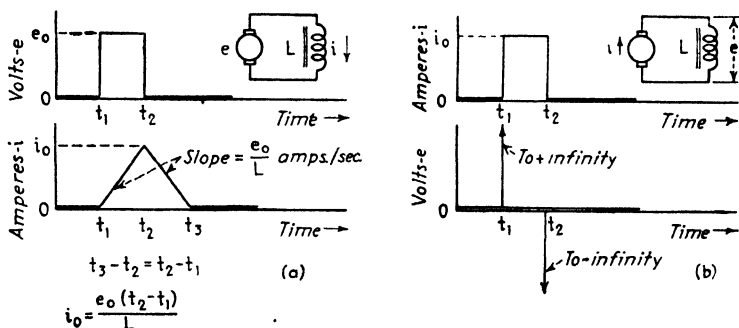



FIG. 42.—Current (*a*) and voltage (*b*) responses to rectangular pulse excitation in an inductive circuit.

current wave whose length at the base is twice the length of the applied voltage pulse.

When a sudden change in current is forced through an inductance, a voltage is built up across the inductance, in accordance with Eq. (32). The more nearly instantaneous the current change (the greater di becomes in a given small time dt), the more nearly infinite the induced voltage is.

Thus the response of an inductance to a rectangular current pulse is two sudden and large changes in voltage, of opposite sign, with zero voltage between (during the pulse). Figure 42 shows the responses of an inductance to voltage and current pulses.

Capacitance.—A pure capacitance C farads stores the energy fed to it in the electrostatic field between the plates of the capacitor.

[1] The integration constant has been omitted, *i.l.*, no initial currents or voltages are assumed in this and the succeeding integrations.

The situation is very similar to that of the inductance, except the current and voltage interchange roles. It is simplest to consider first the response of a capacitance to a current pulse. When a current pulse is forced through a capacitance, the voltage across the capacitance starts to increase at the instant of application and increases linearly with time.

Symbolically,

$$de = \frac{i\,dt}{C} \qquad \text{volts} \tag{34}$$

On integration this becomes for constant i

$$e = \frac{i}{C}t \qquad \text{volts} \tag{35}$$

The similarity of this equation with Eq. (33) should be noted. The current and voltage are interchanged, but otherwise the

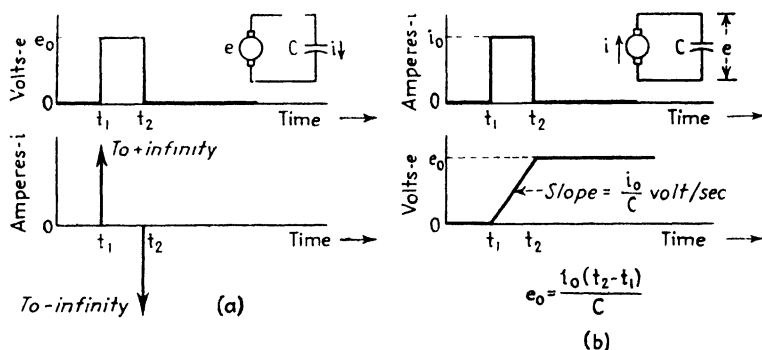



FIG. 43.—Current (a) and voltage (b) responses to rectangular pulse excitation in a capacitive circuit.

relationship is unchanged. Thus the voltage increases linearly with time, at a rate i/C volts per sec, so long as the current persists. At the conclusion of the applied current pulse, the current ceases, and since the generator has infinite internal impedance, the capacitor retains its charge, as shown in Fig. 43.

When a sudden change in voltage is applied, a large current flows through the capacitor. The more nearly instantaneous the sudden change in voltage, the more nearly infinite the current, in accordance with

$$i = C\frac{de}{dt} \qquad \text{amp} \tag{36}$$

Thus the response of a pure capacitance to a rectangular voltage pulse is two sudden and very large changes in current, opposite in sign, with zero current during the pulse. Figure 43 shows the responses of a capacitance to voltage and current pulses.

36. Transient Responses of Pairs of Circuit Elements. Resistance and Inductance in Series.—When R ohms and L henries are *connected in series and e volts suddenly applied across them, the differential equation governing the current, compounded of* Eqs. (31) and (32), is

$$e = iR + \frac{L\,di}{dt} \qquad \text{volts} \tag{37}$$

When this equation is integrated[1] with constant e, it becomes

$$i = \frac{e}{R}(1 - \epsilon^{-Rt/L}) \qquad \text{amp} \tag{38}$$

This equation states that the current is zero at the instant ($t = 0$) of applying the voltage pulse and builds up exponentially, finally reaching a value of e/R amp after a very long time. Numerical substitution reveals that the current reaches a value of $0.63e/R$ amp after L/R sec. The quantity L/R is known as the "time constant" of the circuit.

The physical interpretation of Eq. (38) is not difficult. When the voltage is suddenly applied, the inductance blocks the current flow initially, so that all the voltage is applied across it. The current thereafter builds up in the inductance and would increase linearly were it not for the fact that the current also passes through the resistance across which a voltage drop develops. This voltage drop is subtracted from the voltage applied to the inductance, and in consequence the rate of increase of current falls off. The current continues to increase at an ever decreasing rate until, after a sufficient interval has passed, the current through the inductance ceases to increase and assumes a steady value. There is then no voltage drop across the inductance, and all the applied voltage e appears across the resistance R. This situation agrees with the fact that the final value of current is e/R amp.

[1] Readers not familiar with the solution of differential equations may wish to consult a standard text. The correctness of the integration may be verified by substituting Eq. (38) in Eq. (37), performing the indicated differentiation, and noting that an identity results.

When two equal and opposite sudden changes of voltage (that is, a voltage pulse) are applied to the RL series circuit, two successive transient current responses of opposite sign result. If the length of the pulse is short compared with the time constant L/R, the circuit does not reach equilibrium by the time the pulse ends. The decrease of current that follows has an initial value equal to the final value of the current rise that preceded it The current decrease is expressed by

$$i = \frac{e}{R}(1 - \epsilon^{-Rt_0/L})\epsilon^{-R(t-t_0)/L} \quad \text{amp} \tag{39}$$

where t is measured in seconds from the beginning of the applied pulse and t_0 is the width of the pulse in seconds. This current decrease, like the current increase, is exponential and governed by the time constant L/R sec. After a sufficiently long time (many times L/R sec) the current falls to zero, and the response to the applied voltage is complete.

FIG. 44.- Current responses to a rectangular voltage pulse in an L-R series circuit.

The graphical interpretation of Eqs. (38) and (39) is given in Fig. 44. The superposition of the current rise and fall, which form the current response to the applied pulse, is shown for three cases, in which the pulse width is shorter than, comparable to, and longer than the time constant L/R. It is evident in two of the cases that the current waveform bears little resemblance to the applied voltage.

When a series circuit is excited by a voltage pulse it is of interest to note how the voltage divides across each circuit element. In the RL series circuit, the voltage appearing across the resistance is the current flow, given in Eqs. (38) and (39), times the resistance. The voltage appearing across the inductance is the difference between the applied voltage and the resistance voltage drop.

The response of the RL series circuit to an applied pulse of current is found by inspection of Eq. (37). The first term is

the resulting voltage drop across R and the second term is the voltage across L. At the instant of application of the current pulse, at $t = 0$, the voltage is infinite. This is another way of saying that it is impossible to force a sudden change of current through an inductance, whether or not there is resistance in series with it. In practice the current pulse has a finite initial slope, and the voltage appearing across the series circuit, initially, is given by $e = L\,di/dt$ where di/dt is the initial slope. After the top of the current pulse is reached, the voltage across the inductance becomes zero, and the total voltage drop is that across the resistance. This is the value of the current pulse times the resistance. This value of voltage decays slowly until the current pulse is terminated. The voltage then reverses rapidly and assumes a large negative value, finally reaching zero after an exponential decay.

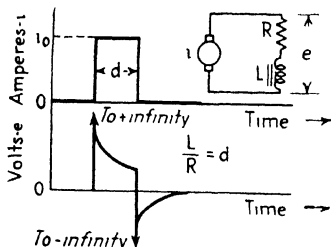



FIG. 45—Voltage response to a current pulse in a series L-R circuit

The voltage drop across an RL series circuit, when a current pulse is applied, is shown in Fig. 45, when the pulse width has a value equal to the time constant L/R.

Capacitance and Resistance in Series.—When C farads and R ohms are connected in series and e volts suddenly applied across them, the integral equation that governs the resulting current is

$$e = iR + \int \frac{i\,dt}{C} \qquad \text{volts} \tag{40}$$

When Eq. (40) is solved, with constant e the solution is

$$i = \frac{e}{R}\epsilon^{-t/RC} \qquad \text{amp} \tag{41}$$

as substitution will prove.

This equation states that when the voltage pulse is first applied, at $t = 0$, the current is initially e/R amp. Thereafter the current decreases exponentially governed by the time constant RC, reaching a value $0.37e/R$ in RC sec. It does not become zero until many times RC sec.[1]

[1] In theory, of course, none of the exponential processes described com-

The physical interpretation of Eq. (41) is as follows: The initial current that flows through the RC series circuit cannot at once establish a voltage across the capacitance. Hence all the applied voltage, initially, appears across the resistance, and the current flow is e/R. As this current continues to flow, the capacitance becomes charged and a corresponding voltage drop appears across it. As the voltage drop across C increases, less voltage is available across R, and therefore the current decreases, the rate of charge of C decreases, and the rate of build-up of voltage across C decreases. In this manner the current decreases at an ever decreasing rate until finally it becomes zero. At this time, no voltage drop appears across R and all the applied voltage appears across C.

FIG. 46.—Current responses to a rectangular voltage pulse in a series R-C circuit.

When the voltage pulse is terminated, the same sequence of events occurs, with opposite sign. The current at the end of the pulse is given by

$$i = \frac{e}{R} \epsilon^{-t_0/RC}(1 - \epsilon^{-(t-t_0)/RC}) \qquad \text{amp} \qquad (42)$$

where t is measured in seconds from the beginning of the pulse and t_0 is the width in seconds of the applied voltage pulse. Figure 46 shows the current response of an RC series circuit to an applied rectangular voltage pulse when the pulse length is shorter than, comparable to, and longer than the time constant RC.

As in the case of the RL series circuit, it is of interest to

pletes itself until an infinite time has passed. In practice, the exponential process is said to be complete when the remaining current or voltage is no longer of practical significance. This happens, in any event, when the remaining current or voltage is smaller than that associated with thermal molecular motion within the conductors present.

examine the voltage drops across R and C when a voltage pulse is applied. The voltage drop across the resistance is, as before, the current in Eqs. (41) and (42) times the resistance. The voltage waveforms across the resistance are similar to the current waves of Fig. 46. The voltage drop across the capacitance is the applied voltage minus the voltage across the resistance.

When a current pulse is applied to the series RC circuit, the voltage drop is given by inspection of Eq. (40). The integral $\int i\,dt$ is zero initially so the initial voltage drop is iR, the peak value of the current pulse times the resistance. As the current pulse persists, the voltage across the capacitor increases linearly, and the net voltage across the circuit increases linearly until the current pulse terminates. The termination of the applied current implies the opening of a switch. Since current does not flow thereafter, the condenser remains charged. If another current pulse of the same sign is applied at a later time, a further increment of charge is added. Figure 47 shows the voltage drop across an RC series circuit.

FIG. 47.—Voltage response to a rectangular current pulse in a series R-C circuit.

Inductance and Capacitance in Series.—When L henries and C farads are connected in series and e volts suddenly applied across them, the differential equation that governs the resulting current is

$$e = L\frac{di}{dt} + \int \frac{i\,dt}{C} \qquad \text{volts} \tag{43}$$

The integral solution of this equation for constant e is

$$i = e\sqrt{\frac{C}{L}}\sin\frac{t}{\sqrt{LC}} \qquad \text{amp} \tag{44}$$

as substitution will reveal.

This equation states that the transient response of the LC series circuit is an oscillatory current of maximum amplitude $e\sqrt{C/L}$ amp and frequency $1/(2\pi\sqrt{LC})$ cycles per sec. Since

there is no dissipative element present, the oscillation continues with undiminished amplitude so long as the pulse lasts. When the voltage pulse is terminated, another oscillation is set up in opposite phase to the first. The oscillation thereafter is the sum of the initial and terminal oscillations, and its amplitude may have any value between zero (pulse width equal to an even number of half cycles) and $2e\sqrt{C/L}$ (pulse width equal to an odd number of half cycles). Figure 48 shows the response.

The LC series circuit is not of much practical importance in pulse circuits because the persistent oscillations do not constitute even an approximate reproduction of the applied pulse. A notable exception is the Guillemin pulse-forming artificial transmission line, discussed in Chap. VI, Sec. 128, which is based on transient analysis of the series LC circuit.

FIG. 48.—Oscillatory response to a rectangular voltage pulse in a series L-C circuit.

37. Transient Response of Pairs of Elements in Shunt.—When a voltage pulse is applied to a pair of elements in a shunt circuit, the current flow to the circuit is simply the sum of the currents in the individual shunt elements, given by Eqs. (31), (33), and (36). The current flow in each element may be of two kinds, a direct flow between terminals, and a circulating flow from one element to the other.

RL in Shunt.—When e volts is suddenly applied across R ohms and L henries in shunt, the current to the circuit is composed of Eq. (31) Plus Eq. (33).

$$i = \frac{e}{R} + \frac{et}{L} = e\left(r + \frac{t}{L}\right) \quad \text{amp} \tag{45}$$

Since the voltage across each shunt element is constrained to be equal to the applied voltage, no circulating current flows between elements. When the applied voltage terminates, at the end of the pulse $t = t_0$ the current thereafter is

$$i = \frac{et_0}{L}\epsilon^{-R_s(t-t_0)/L} \quad \text{amp} \tag{46}$$

until zero amp is reached, where R_s is the (small) resistance of the voltage source. The response is shown in Fig. 49.

When a current pulse is applied, a more complicated situation exists. Initially no current flows through the inductance, and the voltage across it is very large. This voltage, applied across the resistance, causes the resistance to pass current. This is a circulating current. Added to it, in the resistance, is the current of the pulse itself. When the top of the current pulse is reached, there is a large circulating current flowing in the

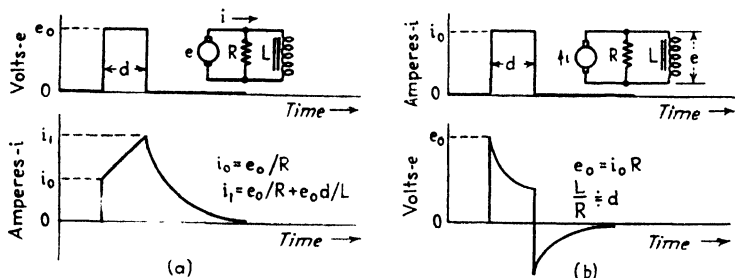



FIG. 49.—Current (*a*) and voltage (*b*) responses to rectangular pulse excitation in shunt L-R circuit.

circuit. The circulating current dies down exponentially in accordance with

$$i_{circ} = \frac{e}{R}\,\epsilon^{-tR/L} \qquad \text{amp} \tag{47}$$

since the circulating current exists in a *series* circuit of L and R. Here e is the voltage across L when the current pulse reaches its full amplitude. As the circulating current decreases the pulse current begins to divide between R and L. Finally the circulating current falls to zero, and all the current passes through the inductance. When the applied current pulse terminates, the net current then existing in the circuit decays exponentially. Figure 49 shows the response of the RL shunt circuit to a current pulse.

RC in Shunt.—When e volts is suddenly applied to R ohms and C farads in shunt, e/R amp flows in the resistance and a brief and very large current flows in the capacitance. As in the RL shunt case, no circulating current can flow. At the end of the applied pulse, the e/R amp becomes a circulating current that discharges the capacitance exponentially.

When a current pulse of amplitude i amp is applied to an RC shunt circuit, no voltage can appear across the circuit initially. The voltage builds up with time across the capacitor, but it does not do so linearly, because as the voltage appears a circulating current can pass through the resistance. Thus the voltage builds up across the circuit exponentially governed by the time constant RC sec. The exponential increase approaches the limit iR volts. When this limit is reached, the condenser is fully charged to this voltage, passes no current, and all the current passes through the resistance. When the current pulse terminates, the condenser discharges exponentially through the

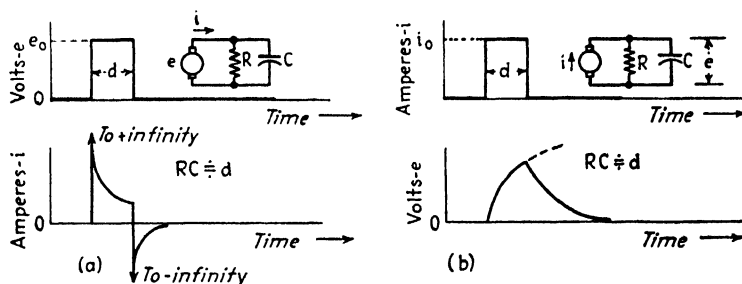



FIG. 50.—Current (a) and voltage (b) responses to rectangular pulse excitation in a shunt R-C circuit.

resistor, the two then being a series circuit through which the circulating current passes. The voltage across the circuit is then equal to the exponentially decreasing current times the resistance. If the current pulse ends before equilibrium is reached, the condenser is not so fully charged and the maximum voltage reached is lower, but otherwise the action of the circuit is similar. Figure 50 gives the current and voltage responses of the circuit.

It should be noted that the voltage across an RC shunt circuit excited with a current pulse is the same as the current in an RL series circuit excited with a voltage pulse, when the time constants are the same. Other similarities are evident in Fig. 51, which is a summary of the pulse responses shown in Figs. 41 to 50 inclusive.

LC in Shunt.—The perfect (resistanceless) shunt-tuned circuit, when excited by either a voltage or a current pulse, sets up a circulating current which tends toward infinite amplitude and

which has the frequency $1/(2\pi\sqrt{LC})$ cycles per second. This circuit, which finds its widest use in r-f and i-f (intermediate-frequency) circuits is treated in Chap. III, Sec. 82.

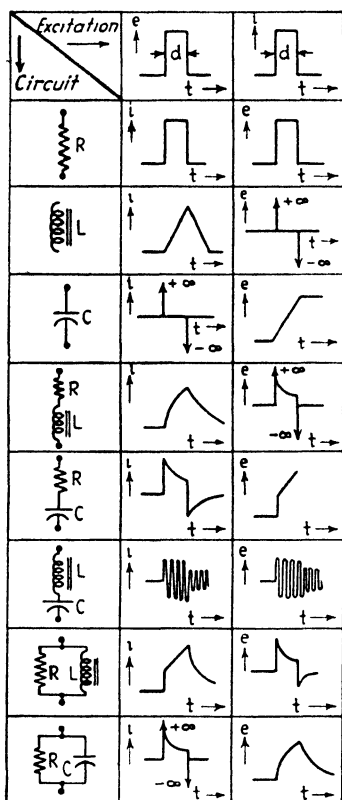



FIG. 51.—Summary of current and voltage responses to rectangular pulse excitation, for simple shunt and series circuits.

38. Circuits of Three or More Elements.—While it is practicable to compute the transient response of series-multiple circuits of three elements by direct integration, the work is tedious and physical interpretation is not simple. Hence the expressions will not be derived here. When the need is met, the three-element circuit will be analyzed by the steady-state method, using the familiar reactance expressions $X_L = \omega L$ and $X_C = 1/\omega C$, where $\omega = 2\pi f$ is the angular frequency. An important example is the real tuned circuit, consisting of L, C, and R. This circuit displays frequency resonance, the sharpness of which varies directly with increasing shunt resistance or inversely with increasing series resistance. In circuits of this type, where the reactances are large compared with the resistance, it is convenient to generalize the circuit action in terms of the figure of merit Q, which is treated at length in Chap. III, Sec. 82.

Another three-element circuit, of great importance in pulse transmission, is the RC shunt circuit with a small series inductance added in the resistance leg. This is the "compensated" coupling connection, commonly used in video amplifiers (pulse amplifiers intended to preserve closely the shape of the waveform amplified). This circuit is analyzed at length by the steady-state method in Sec. 49.

Three-element circuits are often formed by adding a resistance

to the two-element shunt circuits considered in Sec. 37. The resistance is used to develop a voltage from the current flow in the reactive branch of the circuit, when the current waveform has some desired property and must be transformed to a voltage to actuate a vacuum tube. If the added resistance is small, it has negligible effect on the current flow in the circuit, and the small voltage developed may be amplified to whatever level is required.

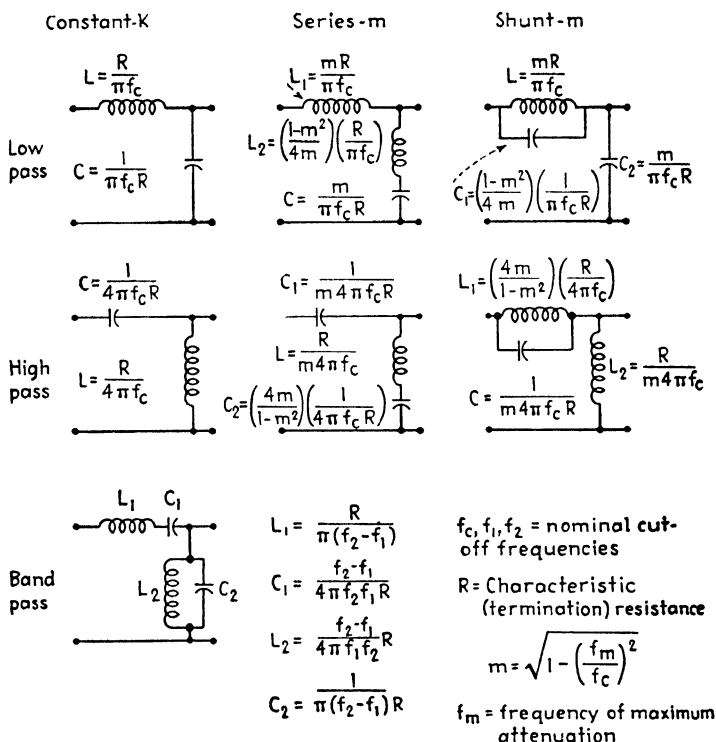



FIG. 52.—Design formulas for constant-K and m-derived filter sections.

Other combinations may be built using the fundamental waveforms shown in Fig. 51. At the present stage little is to be gained from a detailed treatment of these circuits. If the simple relations shown in Fig. 51 are clearly in mind, the transient response of more complicated circuits is often readily apparent, especially if the time constants and impedances of the different parts of the circuit have widely different values, as is often the case.

Finally, a type of multielement circuit of the greatest importance in pulse transmission and generation is the wave filter, particularly of the low-pass and bandpass types. Here again the treatment is most readily understood from the steady-state standpoint. Some familiarity with simple filter design is assumed. Figure 52 will refresh the memory.

39. Transient Response of Vacuum Tubes.—When used in conjunction with the circuits just considered, vacuum tubes have transient characteristics that are of fundamental importance in pulse generation and transmission. The transient response of a vacuum tube is defined in the same way as that of a circuit, that is, as the response to suddenly applied excitation. Since vacuum tubes are voltage-operated devices, only one type of response is considered, the current flow that occurs when a voltage is suddenly applied.

In treating the transient response of vacuum tubes it is impracticable to use parametric terms, such as amplification factor, internal plate resistance, and transconductance, since these quantities are defined for small variations of the tube currents and voltages. Tubes used in pulse work (except in linear video amplification) are customarily driven to saturation or cutoff. Nominal values of the tube constants do not apply to such operation. It is necessary, rather, to employ graphical tube characteristics and to plot thereon the dynamic characteristics over which the tube operates when driven to the limiting values of current and voltage. If great precision is required, this graphical analysis is an involved matter. For engineering purposes, simplified dynamic characteristics may be assumed, which reveal the transient operation and permit reasonably accurate prediction of circuit performance.

There are two basic types of transient tube operation, (1) the grid-voltage cutoff, or negative saturation, and (2) the grid-current cutoff, or positive saturation. These are dynamic responses. Another, passive, aspect of transient tube operation is (3) the effect of the input and output tube capacitances. These capacitances must be made an explicit part of the input and output circuit to determine accurately the transient responses of the connected circuit elements.

40. Grid-voltage Cutoff.—Grid voltage cutoff occurs in a vacuum tube when the total applied voltage between grid and

cathode becomes more negative than the value at which the plate current ceases. So long as the applied grid voltage is more negative than this cutoff value, no plate current flows and the output voltage of the tube circuit is equal to the battery or power-supply voltage. It is evident, therefore, that variations in applied grid voltage, so long as they are more negative than the cutoff value, are not reproduced in the output voltage of the tube. The tube thus *limits* variations in applied voltage when the variations are sufficiently negative to exceed the cutoff value. This limiting action is one of the most important tran-

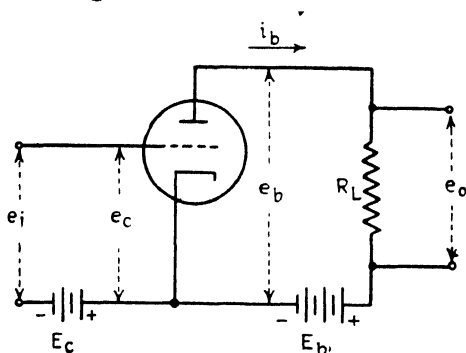



FIG. 53. - Basic circuit for negative grid-voltage cutoff, with fixed bias and resistive load.

sient aspects of vacuum tube operation. It may be used, for example, to produce an approximately rectangular waveform from a sinusoidal wave. Circuits for this purpose are described in Chap. VI. Here we discuss merely the means whereby the dynamic operation of the vacuum tube may be described in terms of the circuit constants and the applied excitation.

The situation is most simple when the tube is operated with fixed grid bias, with fixed battery or power supply voltage in the plate circuit, and when the load impedance is a pure and unvarying resistance. Such a circuit is shown in Fig. 53. In this case, the dynamic characteristic of the tube operation is a straight line (called the "load line") on the plate-voltage plate-current (e_b vs. i_b) static curves, as shown in Fig. 54. The slope of the load line is the inverse of the load resistance, plotted in the same units (for example, milliamperes per volt) as the static characteristics. The load line intersects the plate-voltage scale at the value of the applied battery or power supply voltage. The

intersection P of the load line with the value of fixed grid bias is the operating point of the circuit. Positive applied grid voltage causes the values of e_b and i_b to assume values on the load line increasing upward to the left, whereas negative applied

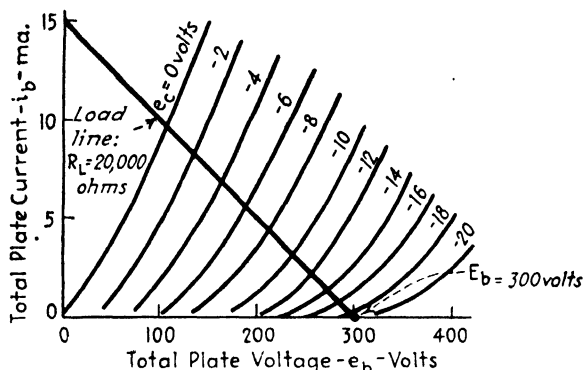



FIG. 54.—Load line method of determining dynamic characteristic of amplifier.

grid voltage causes these values to move downward to the right along the load line. The transient operation of this circuit may be investigated directly on the load line by choosing successive values of e_c, the total applied grid voltage, and reading, from the load line, the corresponding values of plate current. The product

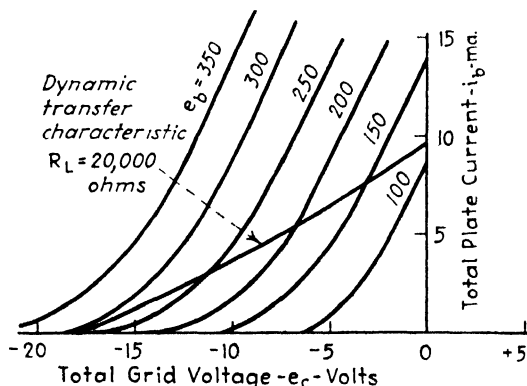



FIG. 55.—Transfer of plate load line to grid-voltage ordinates.

of the plate current times the load resistance gives the output voltage.

The dynamic action of the circuit can be seen more clearly if the load line is transformed to the plate-current vs. grid-voltage (i_b vs. e_c) static curves as shown in Fig. 55. The load line then

becomes the dynamic *transfer characteristic* between applied grid voltage and plate current. In general this transfer characteristic is not a straight line, but is curved by virtue of the nonlinear relation between grid voltage and plate voltage for a given plate current. The transfer characteristic is readily plotted by transposing, point by point, corresponding values of e_c, e_b, and i_b from Fig. 54 to Fig. 55.

When a negative rectangular voltage pulse is applied to the grid circuit, a rectangular pulse of current results in the direction

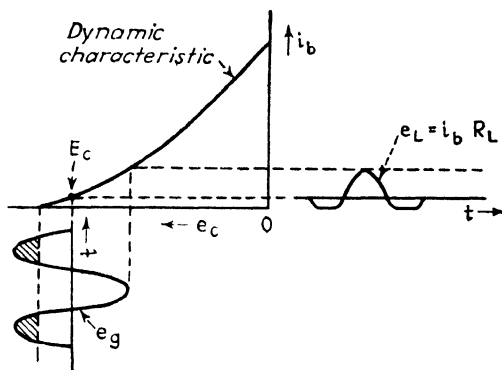



FIG. 56.—Negative limiting by grid-voltage cutoff.

of decreasing current. Any variation in the top of the applied pulse is smoothed off, if it lies to the left of the cutoff value as shown. Moreover, the curvature of the transfer characteristic does not introduce any curvature in a rectangular pulse. The curvature merely governs the amplitude of the plate-current pulse relative to the cutoff grid voltage and the fixed grid bias. When a nonrectangular pulse is applied, on the other hand, variations in the slopes of the pulse below the cutoff value are reproduced and are modified by the curvature of the transfer characteristic, as shown in Fig. 56.

A complication ensues when, as is often the case in practice. the tube is self-biased. In this case the grid-bias voltage is obtained from an *RC* shunt circuit in the cathode, which produces a bias voltage proportional to the sum of the average plate current and the average grid current. The value of the *RC* time constant is usually very long compared with the pulse width. When a series of pulses is applied in the negative

direction, the average plate current decreases, and the grid-bias voltage changes as a result of the excitation. Thereafter, so long as the pulse amplitude, shape, and repetition frequency (pulse rate) remain the same, the grid bias remains in equilibrium.

When self-bias is employed, the load line and transfer characteristics are drawn as with fixed bias (Figs. 54 and 55) except that the value of E_c is that corresponding to the average plate current when excitation is applied. If the average plate current can be measured under excitation conditions, there is no further difficulty. If it must be computed, some difficulty ensues because the average plate current cannot be determined until the plate-current pulse is plotted, and this cannot be done until the appropriate value of grid bias is known. In such a case, the graphical process is performed assuming fixed bias (at the nominal value specified by the manufacturer) to determine the approximate amplitude and shape of the plate-current pulse. The amplitude of the plate-current pulse is then averaged over the pulse interval and the average value subtracted from the nominal (fixed bias) value to obtain the effective value of the plate current under excitation. This effective value is then multiplied by the cathode resistance to determine the actual grid-bias value. If this value departs markedly from the fixed bias previously assumed, the operating point is shifted along the load line to the actual value of grid bias so computed. The load line then approximates the operating conditions so closely that further refinement seldom serves any purpose, although the true operating point may be approached as closely as desired by successive averaging of assumed plate-current pulses.

Further difficulties appear when the load impedance is not a pure and unvarying resistance. The load impedance may be, for example, any one of the series and shunt circuits shown in Fig. 51, all of which contain reactive elements. The load line is then no longer a straight line, but a combination of two or more curved lines, representing the corresponding values of voltage and current across the load impedance. When the grid voltage applied is sinusoidal, the dynamic characteristic is generally a distorted ellipse that may be approximated by a straight line. But when the applied grid voltage changes suddenly, as in a rectangular pulse, the situation is far from simple. In cases where the load impedance is predominantly resistive, it is often

possible to neglect the reactance and proceed as in Figs. 54 or 56. But this assumption cannot be entertained when the pulse is narrow and has very steeply rising sides, because the shunt capacitance inevitably associated with tubes and wiring prevents the voltage in the output circuit from changing instantaneously. Care must be taken in computing the transient response of tube circuits if the load impedance is a complex combination of reactive and resistive elements. As stated previously, it may then be simpler to resort to the steady-state analysis (Sec. 44).

When, as in the cases illustrated in Fig. 51, the current and voltage waveforms in the load impedance differ markedly, the

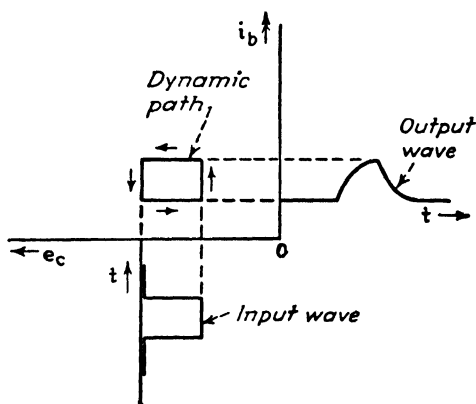



FIG. 57.—Dynamic action of pulse-excited amplifier with reactive load.

dynamic characteristic is not a straight line. It is an odd-shaped figure on which the points represent corresponding values of current and voltage in the impedance. If, for example, the input voltage wave has a flat top (Fig. 57), all the corresponding values of current lie on the same voltage ordinate and that part of the dynamic characteristic is a vertical straight line. Such a line has no evident use in the load line method previously outlined. In such cases, experimental study of the circuit with an oscilloscope is indicated.

For computational purposes, it is often permissible to assume that the vacuum tube is a current generator which, when excited by a rectangular pulse of voltage in the grid circuit, applies a rectangular pulse of current to the load impedance. The voltage waveforms of Fig. 51 then describe the output voltage. If the internal plate resistance r_p of the tube is large compared with

the resistive portion of the load impedance (series) this approximation is often valid. But graphical analysis of transient tube behavior is usually dispensed with when large reactive components are present in the load impedance, in favor of oscilloscopic study.

41. Grid-current Cutoff.—The limiting action described in the preceding section occurs only when the pulse is applied in negative polarity. When positive pulses are applied the plate current may be limited by the effect of grid current. If the grid circuit has poor regulation (that is, high series resistance), the grid

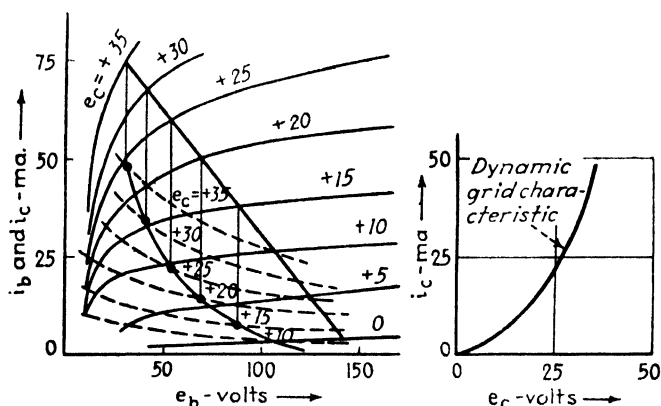



FIG. 58.—Development of dynamic grid characteristic from plate load line and i_b-i_c curves.

current will prevent the grid voltage from rising rapidly and a corresponding limit is placed on the plate current. If the input circuit has good regulation, the regulation may be lowered purposely by inserting a resistance in series with the grid.

The limiting action in this case occurs wholly in the grid circuit, including the grid and cathode of the tube. However, the value of grid current that flows in response to positive grid voltage depends in part on the plate voltage, which is varying along the load line. Here, again, is a complicated situation that yields most readily to the experimental approach, that is, observation of the waveforms with an oscilloscope.

If graphical computation is required or preferred, it is necessary to determine the dynamic characteristics relating total applied positive grid voltage e_c and resulting total grid current i_c. This characteristic is determined from the combined i_b-vs.-e_b

and i_c-vs.-e_b curves, such as are shown in Fig. 58. These curves, it should be mentioned, are not generally published by tube manufacturers but may be measured readily. The plate-circuit load line is erected on the i_b-vs.-e_b curves in the usual manner and corresponding values of e_c and e_b obtained from points on the load line. These values of e_c and e_b are then applied to the i_c-vs.-e_b curves to obtain the corresponding values of i_c. The i_c values are then plotted against the e_c values, as shown in Fig. 58.

The effective value of total applied grid voltage e_c is obtained by erecting a grid-circuit load line on the dynamic grid charac-

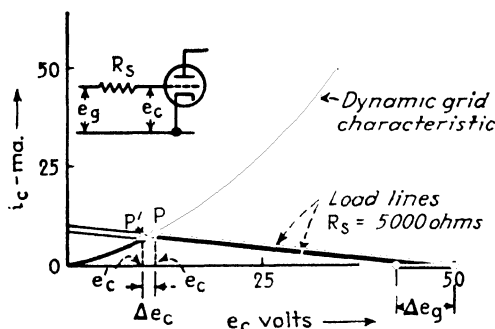



FIG. 59.—Dynamic action of positive limiting due to flow of grid current.

teristic, of slope equal to the inverse of the resistance in series with the grid. This line intersects the voltage axis at the value of applied grid voltage (bias amplitude in the absence of the applied voltage pulse). The corresponding value of voltage across the tube e_c' is given at the intersection P' (Fig. 59). When the voltage pulse is applied, e_g increases proportionately in the positive direction, and e_c also increases as the intersection moves to P. If the series resistance is large, the load line is nearly horizontal and a large change may occur in e_g with a very small corresponding change in e_c. This is the limiting action due to the grid current flow. The limiting is sharper as the slope of the dynamic grid characteristic increases.

When the magnitudes of e_c and e_c' are determined, reference to the dynamic plate characteristic (at left in Fig. 58) gives the corresponding values of plate current i_b and i_b', from which the output voltage may be computed.

If the applied pulse is not rectangular, points on the applied waveform may be selected at equal intervals of time, and cor-

responding positions of the grid load line assumed for the corresponding values of e_g. When the process is carried through to the corresponding values of plate current, these may be plotted at the stated intervals of time to obtain the plate-current waveform. This method of computing grid current limiting may be used only when the plate load is a pure and constant resistance for the reasons previously outlined.

It should be remarked that the positive limiting action of grid current is never as sharp or as complete as the negative limiting caused by grid-voltage cutoff. In the grid current case some additional grid voltage e_c is always developed as the applied signal is increased in the positive direction. Negative saturation, on the other hand, reduces the plate current to zero (in sharp-cutoff tubes at least) and thereafter no further change in grid voltage in the negative direction can affect the plate current.

42. Transient Effect of Tube Capacitances.—The input and output capacitances of a vacuum tube have a decided effect on its transient behavior whenever they constitute an appreciable fraction of the shunt capacitance connected to the tube. In pulse circuits shunt capacitance external to the tube is avoided whenever the tube is required to reproduce rapid changes in voltage. In this case the tube capacitances, together with wiring capacitance, become the limiting factors.

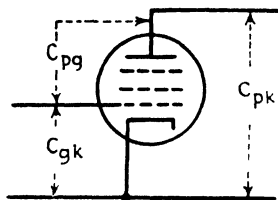



FIG. 60.—Interelectrode capacitances

The output capacitance of a tetrode or pentode vacuum tube consists essentially of the plate-cathode capacitance, C_{pk}, as measured across these terminals with the tube in a "cold" or nonoperating condition. This is the case because the grid-to-plate capacitance C_{gp} is very small compared with the plate-to-cathode value. In triode tubes, (Fig. 60) the output capacitance consists of C_{pk} with C_{pg} and C_{gk}' in series across it. C_{gk}' is the effective capacitance, including external circuit capacitance, measured between grid and cathode of the tube. Typical values of output capacitance in receiving tubes are 2 to 5 $\mu\mu$f in triodes and 5 to 10 $\mu\mu$f in pentodes.

The input capacitance is a dynamic quantity that increases with the plate current. The dependence on plate current arises

from the fact that the reactive current entering the grid circuit is composed in part of current to the plate through the grid-plate capacitance. The current passed by C_{gp} depends upon the potential of the plate which in turn depends upon the flow of plate current through the load impedance. When the load impedance is a pure resistance R_L, the input capacitance of the tube is approximately

$$C_{in} = C_{gk} + C_{gp}\left(1 + \frac{\mu R_L}{r_p + R_L}\right) \tag{48}$$

where r_p is the internal plate resistance and μ the amplification factor of the tube. When the load resistance is large compared with r_p, this equation reduces to the more familiar

$$C_{in} = C_{gk} + C_{gp}(1 + \mu) \tag{49}$$

The input capacitance should be measured with the tube in normal operating condition.

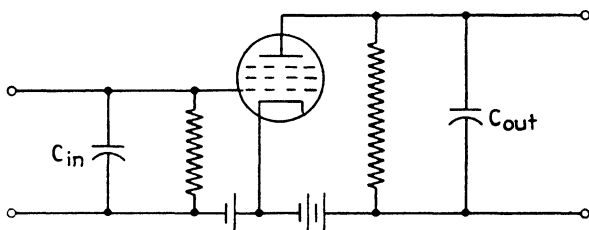



FIG. 61.—Shunt capacitances at input and output of amplifier stage.

The effect of the input and output capacitances on tube operation is the same as that of any external capacitance in shunt, that is, the capacitance prevents any sudden change in the voltage across the circuit except at the expense of a heavy current flow. To illustrate the effect, consider Fig. 61. The shunt resistances in grid and plate are considered to be so large that negligible current flows through them. When a voltage pulse is applied to the grid, the voltage does not immediately arise across the grid circuit. Rather the rise in voltage is linear with time, in accordance with Eq. (35). The rate of the rise is proportional to the current supplied by the grid circuit and inversely proportional to the shunt capacitance.

If the grid circuit cannot supply a heavy current during the pulse and if the shunt capacitance C_{in} is large, the rate of rise of voltage across the grid and cathode is relatively slow, and

it may be a poor reproduction of the leading edge of the applied pulse. Suppose the input capacitance is 10 $\mu\mu f$, and the maximum value of the current pulse delivered by the input circuit is 1 ma. Then, by Eq. (35) the maximum rate of rise of voltage at the grid is

$$\frac{e}{t} = \frac{i}{C} = \frac{10^{-3}}{10^{-11}} = 10^{8} \qquad \text{volts per sec}$$

or 100 volts per μsec. If the leading edge of the pulse completes itself in 0.1 μsec (as is common in practice) the maximum response of the grid voltage in that time is 10 volts, which may be far below the maximum value of the applied pulse.

The speed of response is increased if the resistance value is reduced below the near-infinite value considered above. The response is then governed by the RC time constant, as given by Fig. 50. But the effect of the capacitance can never be completely overcome, since lowering the resistance also lowers the voltage that can be built up across it by a given current source.

A similar situation exists in the plate circuit. Here the output capacitance governs. The current pulse is supplied by the plate current of the tube, which is often larger than that available in the grid circuit. Hence a more rapid rise of voltage is possible. If the peak value of plate current is 100 ma (as it may be even in small receiving tubes when operated in pulsed circuits) and the output capacitance is 10 $\mu\mu$f, the rate of voltage rise is 10,000 volts per μsec, and a 100 volt amplitude may be reached in 0.01 μsec. In practice such rapid rates of rise are not often required. But on the other hand the total value of shunt capacitance (output capacitance, capacitance of circuit elements and wiring, input capacitance of the following stage, etc.) is usually considerably larger than 10 $\mu\mu$f.

The effect of shunt capacitance is one of the basic limitations in pulse circuit design, and frequent reference to it will be made in the following pages, particularly in the discussion of video amplifiers (Sec. 49).

Effect of Transit Time.—The time taken by the electrons to pass from cathode to grid in a vacuum tube (transit time) has an important effect when the rise and fall of the applied voltage occurs in comparable time. This effect is prominent at radio frequencies above 500 megacycles, since the transit time of

conventional tubes under typical operating conditions is of the order of 10^{-9} sec. In pulse work, however, pulses shorter than 10^{-7} sec are not generally employed, so transit time need not be considered.

43. Transient Operation of Diodes.—The diode vacuum tube has two attributes in transient operation. It can act as a limiter of negative or positive pulses, and it can develop a direct voltage very nearly equal to the peak value of a series of applied pulses.

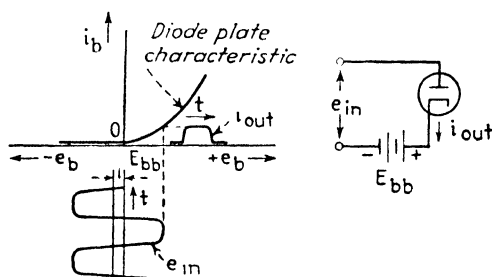



FIG. 62.—Negative limiting in a diode circuit.

The negative limiting action of a diode, shown in Fig. 62, depends on the fact that the diode ceases to conduct when its anode voltage is driven negative with respect to the cathode. Consequently when a properly poled bias voltage is provided in series with the applied voltage, conduction continues so long as the bias voltage is not exceeded by applied voltage of

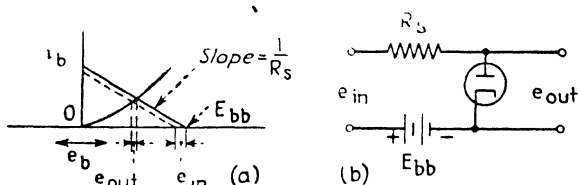



FIG. 63.—Positive limiting in a diode circuit.

negative sign. Limiting action occurs at values of the applied voltage greater (more negative) than this limit.

The positive limiting action of the diode is very similar to that due to grid current in a multielement vacuum tube. The i_c-vs.-e_c characteristic shown in Fig. 58 is replaced by the static i_b-vs.-e_b curve of the diode, and the load line is constructed in the usual way. The construction of Fig. 59 then indicates the extent of the limiting action on the applied positive pulse. Here

again a bias voltage is customarily used to provide a region within which no limiting occurs. The bias voltage may also be connected in the opposite polarity to that shown in Fig. 63. No limiting action occurs so long as the applied voltage does not exceed the bias value. At higher values of applied voltage, conduction occurs and the limiting action takes place. As in the case of grid-current limiting, high series resistance is required in the circuit supplying the pulse to obtain a flat limiting characteristic.

The third aspect of diode operation, developing a direct voltage equal to the peak value of a series of pulses, is of interest

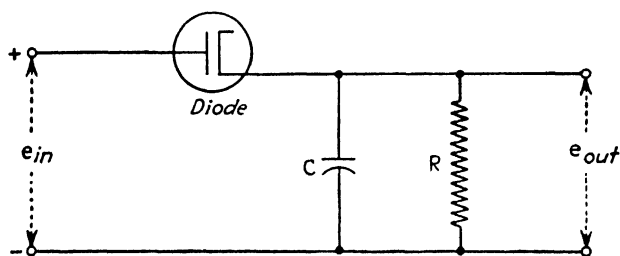



FIG. 64.—Diode detector for following variations in pulse amplitude.

whenever it is necessary to develop a control voltage from a series of pulses or when it is necessary to establish the peak value of a series of pulses against a fixed reference. The latter function is often called "d-c insertion."

The principle involved is that of the peak diode detector. The basic circuit, shown in Fig. 64, is a diode in series with a capacitor. The applied pulses are connected in such polarity that the peak value causes conduction (positive voltage on the anode with respect to the cathode). The conduction current is accumulated by the capacitor. So long as the peak value of the applied pulse exceeds the voltage on the capacitor, current flows and the capacitor charge increases. When the voltage across the capacitor equals the peak value of the applied voltage, no net voltage is present across the diode during the pulse and no further current flows. If the capacitor and diode have negligible leakage, equilibrium is reached at this point. Any leakage present is made good by periodic recharging of the capacitor from the pulse source.

When it is desired that the voltage across the capacitor should

follow changes in the pulse amplitude, which is precisely the operation of a diode detector, resistance is inserted to allow the voltage across the capacitor to decrease when the pulse amplitude falls below the charging level.

The diode d-c level circuit is used in automatic sensitivity control circuits and in automatic brightness control of cathode-ray indicators. In more elaborate form, known as the "clamping circuit" (Sec. 144, Chap. VI), it is used to hold the minimum value of an indicator deflection waveform at a fixed value regardless of changes in the amplitude of deflection. The fixed value of deflection may, for example, correspond to the center of a ppi indicator.

44. The Steady-state Analysis of Pulse Circuits.—Earlier in this chapter it was pointed out that the transient analysis of circuit behavior is particularly suitable when pulses are transmitted, but that the evaluation of transient behavior may be difficult in complex circuits. Then it is necessary to fall back on the steady-state analysis, which, while less direct than the transient, often yields results with less labor. The steady-state method is particularly appropriate when we consider the capabilities of an entire transmission system or network. Steady-state measurements made at the input and output terminals of the network will, with the proper analysis, provide information on the distortions undergone by the pulse in passing through the network. The same information may be obtained directly by applying the pulse to the input and observing the output on an oscilloscope. But the steady-state analysis gives much more general information than is revealed in particular measurements. In any event, the steady-state approach should be thoroughly understood by those who design radar systems.

In the steady-state method, the basic entity is a sinusoidal wave of current or voltage, which is applied to a network. The amplitude and phase angle of the resulting response (voltage or current, respectively) is computed or measured. The process is repeated for as many different frequencies of applied excitation as are necessary to the problem, the amplitude and phase of the response being obtained for each applied frequency. This information is assembled in two *system functions*, the amplitude system function and the system phase function. These are simply characteristic curves, amplitude vs. frequency and phase

vs. frequency, which describes the response of the transmission system or network.

When the system functions are available, the response of the system to a particular waveform is found by breaking down the applied waveform into its components (its fundamental and harmonic frequencies). By means of the Fourier series or integral this process may be carried out for any waveform of practical significance. The results of the waveform analysis are the so-called "spectra" of the waveform, the "amplitude spectrum" and the 'phase spectrum." These spectra describe the amplitudes and phases of a group of sinusoidal components that, when added together, generate the applied waveform.

To obtain the waveform at the output of the transmission system (the "output pulse") the spectra of the applied waveform are combined with the system functions. The amplitude spectrum is multiplied by the amplitude system function, and the phase spectrum is added to the phase system function. The multiplication and addition give rise to two new spectra, those corresponding to the output waveform. These spectra are then synthesized, that is, the sinusoidal components described by the spectra are added together. The result is the waveform of the pulse as it appears at the output of the network.

The output pulse may then be compared directly, in size, shape, and time of occurrence, with the applied pulse. The difference in shape represents the waveform distortion introduced by the transmission system, and the time difference, between application of the input pulse and appearance of the output pulse, represents the delay of the transmission system.

As we shall prove later, waveform distortion, which is usually of primary importance, can be minimized by employing transmission systems whose system functions meet certain requirements: in particular the amplitude system function must be substantially a constant for all frequencies in the spectrum of the applied pulse, and the phase system function must be directly proportional to frequency (time delay equal at all frequencies) over the same region. In other words, the transmission system must display no discrimination in amplitude among the several component frequencies of the pulse, and it must delay all such frequencies an equal amount of time. If these requirements are met, the output waveform has the same

shape as the input waveform, the only changes being in amplitude and time of occurrence.

The problem of steady-state analysis and synthesis thus breaks down into four basic steps:

1. Determination of the spectra of the input (applied) waveform
2. Determination of the system functions
3. Combination of spectra and system functions
4. Synthesis of the output waveform

These steps are outlined in detail in the following paragraphs.

45. Determination of Pulse Spectra.—It is well appreciated in communications that any waveform (excepting mathematical oddities that do not occur in practice) may be represented as the sum of sinusoidal and cosinusoidal components with appropriate amplitudes and phases. This is a statement of the Fourier theorem. If the waveform is repetitive (repeats itself regularly over an indefinitely long time), the components are discrete and may be represented as terms in a sum called the "Fourier series." If the waveform is transient, that is, if it occurs once and is nonrepetitive, the sinusoidal components form a continuous spectrum that is represented by the Fourier integral.

The Fourier series of a repetitive waveform $W(t)$ is

$$W(t) = \frac{a_0}{2} + a_1 \cos 2\pi ft + a_2 \cos 2(2\pi f)t + a_3 \cos 3(2\pi f)t + \cdots + b_1 \sin 2\pi ft + b_2 \sin 2(2\pi f)t + b_3 \sin 3(2\pi f)t + \cdots \quad (50)$$

Written in more compact form

$$W(t) = \frac{a_0}{2} + \sum_{k=\infty}^{k=1} a_k \cos k2\pi ft + b_k \sin k2\pi ft \quad (51)$$

These equations state that the waveform $W(t)$ may be represented as the sum of an infinite number of cosinusoidal components of amplitude a_k and sinusoidal components of amplitude b_k, where k is the order of the harmonic, 1 for the fundamental, 2 for the second harmonic, 3 for the third harmonic, and so on to infinity.

It is thus evident that the spectra of a waveform may cover an infinite frequency range. In practice, of course, the frequency

range is always finite, and this is accounted for in the series by the fact that the amplitudes a_k and b_k are all zero outside the finite frequency range. The fundamental frequency of the waveform is f (angular frequency $2\pi f$), and the harmonics are whole multiples of the fundamental value.

The roots of the Fourier analysis are the amplitudes of the harmonics a_k and b_k. These are determined by the size and shape of the applied waveform. They are computed by the following integrals:

$$a_k = 2f \int_{-1/2f}^{1/2f} W(t) \cos k2\pi ft \, dt \tag{52}$$

$$b_k = 2f \int_{-1/2f}^{1/2f} W(t) \sin k2\pi ft \, dt \tag{53}$$

These equations show that the amplitude of a particular cosinusoidal harmonic is found by multiplying the waveform by the cosine of that frequency and integrating the result over the fundamental period of the wave $1/f$. The amplitude of a sinusoidal harmonic is found the same way, except that the waveform is multiplied by the *sine* of the same frequency. Whether or not the integration can be performed simply depends entirely on the waveform $W(t)$. In many cases, the resulting product is an integrand that may be found in tables of integrals. In other cases, the integration must be performed graphically or mechanically or by series methods.

When the amplitudes a_k and b_k have been found in this manner, they are combined by the following relationships:

$$c_k = \sqrt{a_k^2 + b_k^2} \tag{54}$$

and

$$\theta_k = \tan^{-1} \frac{b_k}{a_k} \tag{55}$$

Here c_k is the amplitude and θ_k is the phase angle of the kth harmonic of the fundamental. When all the harmonics thus represented are added, with due regard to sign, on the same amplitude scale and time scale, the original waveform $W(t)$ results. Hence c_k and θ_k represent, respectively, the amplitude spectrum and the phase spectrum of the pulse waveform. The process of computing the spectra, while apparently cumbersome at first glance, is not difficult if the integrations of Eqs. (52) and (53) can be performed readily.

46. Spectra of the Repetitive Rectangular Pulse.—A simple and appropriate example of the foregoing process is the spectra of the repetitive rectangular pulse, shown in Fig. 65. We can simplify the computation considerably by placing the waveform symmetrically about the zero time axis, as shown. In this case, the function is known as an "even" function, since it has the same value at corresponding positive and negative values of t.

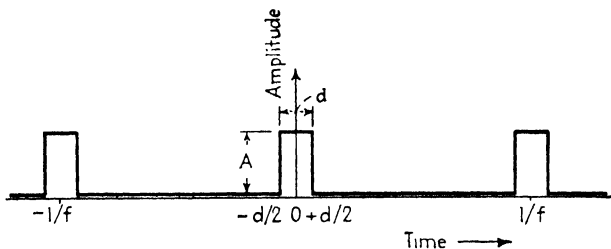



FIG. 65.—Basic repetitive rectangular pulse, arranged as a even function (mirror symmetry about zero time axis).

When an even function is integrated in Eq. (53), it is found that the integral b_k is zero.[1]

Then the spectra become

$$c_k = a_k \tag{56}$$

and

$$\theta_k = \tan^{-1} 0 = 0 \qquad \text{deg} \tag{57}$$

In view of Eq. (56) we find the amplitude spectrum c_k of the rectangular pulse from Eq. (52), the integral for a_k. As shown in Fig. 65, the pulse amplitude is A volts or amp, its width d sec, and its fundamental frequency f cps. The limits of the integral are $+d/2$ and $-d/2$, since the pulse amplitude is zero outside these limits throughout the period $1/f$. Then Eq. (52)

[1] The reason for this property is not difficulty to discover. The sine factor in Eq. (53) has positive values on one side of the zero time axis and negative values on the other. When multiplied by a symmetrical function $W(t)$, the product comprises a given positive area on one side of the axis and an equal negative area on the other side. The sum of these two areas (the integral) must be zero. Conversely if the waveform displays skew symmetry about the zero axis (amplitudes equal but opposite in sign at corresponding negative and positive values of t), the amplitudes a_k all become zero, by similar reasoning. The mirror symmetry of Fig. 65 is often present in radar pulses, but the skew symmetry is present (Fig. 66) less often.

becomes

$$a_k = 2f \int_{-d/2}^{d/2} A \cos k2\pi ft \, dt \tag{58}$$

This is evidently a simple integral. Reference to standard tables reveals that it is equal to

$$a_k = 2Adf \frac{\sin k2\pi fd/2}{k2\pi fd/2} \tag{59}$$

From this expression we can plot the amplitudes $a_k = c_k$ as a function of the frequency and thus obtain the amplitude spectrum of the pulse. First let us examine the term $a_0/2$, the first term in the series of Eq. (50). This is the average value or d-c component of the wave. Substituting $k = 0$ in Eq. (59), we obtain

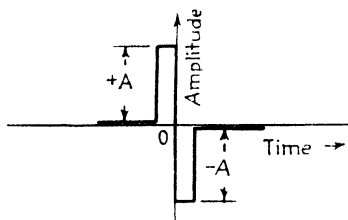



FIG. 66.—Skew symmetrical pulse, an odd function.

$$\frac{a_0}{2} = Adf \frac{\sin 0}{0} = Adf \tag{60}$$

Thus the average value of the wave is equal to the peak value A times the duty cycle df (pulse width times pulse rate, as defined in Chap. I).

The fundamental $k = 1$ and harmonic frequencies $k = 2, 3, 4,$ etc., of the pulse spectrum are determined by substituting the appropriate values of k, f, and d in Eq. (59).

Figure 67 is a plot of Eq. (59), amplitude vs. frequency. The frequency is taken in units of kf, the fundamental frequency f times the harmonic order k. As kf increases, a_k decreases since the sine of a given value decreases more rapidly than the given value itself. When $kf = 1/d$, $\sin k2\pi fd/2$ becomes $\sin \pi = 0$. Hence the amplitude of the corresponding harmonic is zero. This is the $1/fd$ th harmonic, and the frequency at which it occurs is $1/d$ cps, or the inverse of the pulse width. This value of frequency is known as the "first zero" of the pulse spectrum.

As the order of the harmonic k increases above $1/fd$, the amplitudes of the harmonics increase and decrease, with negative sign, until the frequency $kf = 2/d$ is reached. At this point $\sin k2\pi fd/2 = \sin 2\pi = 0$. This is the second zero of the spectrum. Thus the amplitudes of the harmonics rise and fall,

with alternation of sign, as the frequency is increased. The zeroes of the spectrum occur at frequency intervals equal to the inverse of the pulse width. Figure 67 shows the spectrum plotted out to the fourth zero.

It should be noted that the spectrum of the rectangular pulse is theoretically of infinite extent, since the harmonics never completely disappear, except at the "zeroes", no matter how high the frequency considered. To obtain true reproduction of a rectangular pulse, therefore, the transmission system must

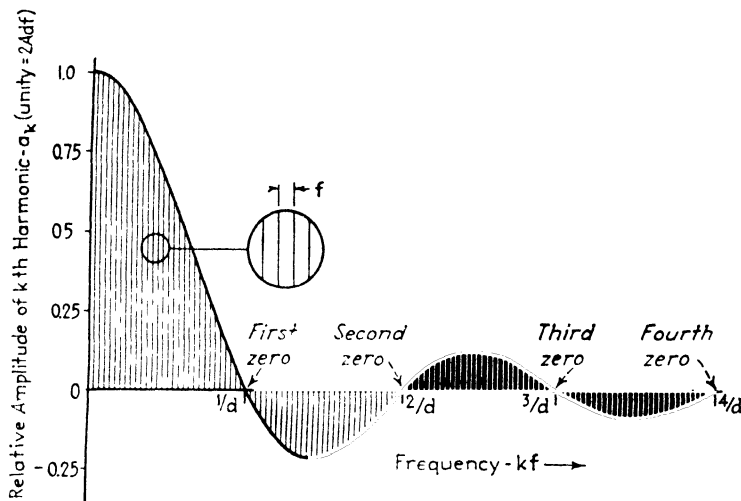



FIG. 67.—Amplitude spectrum of the repetitive rectangular pulse, composed of discrete harmonics.

respond without discrimination in amplitude or time delay over an infinite bandwidth. Since this is impossible in practice, it is impossible to transmit a rectangular pulse without some distortion. But if the bandwidth of the system extends out to the fourth or fifth zero of the spectrum quite acceptable reproduction will result.

It will be noted in Fig. 67 that the harmonics of greatest amplitude occur between the fundamental frequency f and the first zero frequency $1/d$. This is evidence of the fact that the most important region of the pulse spectrum lies between these limits. A general rule, in transmitting pulses of any shape, is that the transmission bandwidth should extend from the pulse-repetition frequency at the lower limit to the inverse of the pulse

width at the upper limit. The distortion introduced by such a system, as we shall see in detail in Sec. 54, is appreciable, but can be tolerated in many pulse systems.

As a numerical example, consider a pulse of width $d = 1$ μsec and peak amplitude $A = 10$ volts transmitted at a rate of $f = 500$ pulses per sec. The average value of the pulse is

$$Adf = 10 \times 10^{-6} \times 500 = 0.005 \text{ volt.}$$

The amplitude of the fundamental component is $2Adf = 0.01$ volt, and this occurs at 500 cps. The amplitude of the harmonics decreases at higher frequencies and reaches its first zero at a frequency of $1/d = 1{,}000{,}000$ cps. This first zero occurs at the $1/fd = 2{,}000$th harmonic of the fundamental.

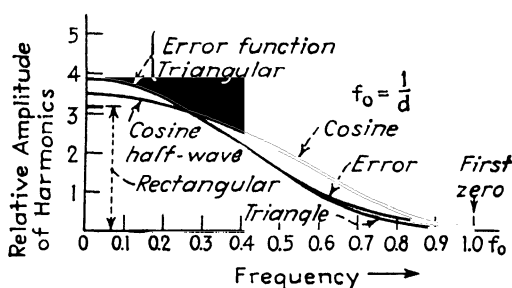



FIG. 68.—Amplitude spectra of cosine, triangular, and error function pulses of equal energy, plotted to first zero. (*After J. A. Stratton.*)

In such a pulse there are a vast number (several thousand) significant harmonics. It must be remembered that the spectrum is nevertheless discontinuous. If the spectrum is explored by a spectrum analyzer, no components will be found except at frequencies of 500, 1,000, 1,500 and successive multiples of 500 cps. To illustrate this fact, a magnified portion of the spectrum is shown in Fig. 67.

The phase spectrum of the repetitive rectangular pulse is simple, as it is in any function having mirror symmetry and centered on the zero axis,

$$\theta_k = 0 \tag{57}$$

This equation states simply that the phase angle of all harmonics is the same at the start of the pulse, for example, zero and increasing. The fact that the phase spectrum happens to be zero does not make it any simpler for the transmission system to avoid phase distortion, because, as we have seen, the phase

system function is added to the phase spectrum of the pulse (zero in this case). Thus in the case of an even function, the phase spectrum of the output pulse is equal to the phase system function, and any phase irregularities in the transmission system are transposed directly to the spectrum of the output pulse.

The rectangular pulse is an important example because many of the pulses employed in radar have substantially rectangular form. The spectra of other typical pulse shapes, plotted out to the first zero, are shown in Fig. 68. These shapes are among the few which produce integrals for which fairly simple analytic solutions are available. It is evident that the amplitude spectra of dissimilar pulse shapes are remarkably similar. It should be noted that the pulse width of the nonrectangular pulses is taken at the half-amplitude point and that the energy (the integral of the amplitude squared with respect to time) is the same in each pulse. The phase spectra of these pulses are zero, since all are even functions.

47. The Fourier Integral.—The foregoing analysis, based on the Fourier series, applies only when the pulse is repetitive. Although the pulses employed in radar are ordinarily sufficiently repetitive to permit the series formulation, it is worth while to extend the concept to the case of a single, nonrepetitive transient pulse.

When a transient pulse occurs, there is no pulse rate f in evidence and the series of Eq. (50) has no meaning. We may extend the concept of the series, however, by allowing the pulse rate to become lower and lower, until finally the interval between pulses is infinite. By this reasoning the pulse rate becomes zero, and the k, order of the harmonic at any frequency in the spectrum, becomes infinite. Thus the product kf remains finite. In the limit of the transient repetitive pulse kf has the value

$$kf = \frac{\omega}{2\pi} \tag{61}$$

Here ω is a new angular frequency, which is the frequency scale of the pulse spectrum. We have in this limit an infinite number of infinitesimally spaced harmonic components, that is, a continuous spectrum. If $S(\omega)$ is a complex quantity that represents the amplitude and phase spectra as functions of ω, then the

waveform $W(t)$ is the integral of all the harmonic components. In symbolic form,

$$W(t) = \int_{-\infty}^{+\infty} S(\omega)\epsilon^{j\omega t}\, d\omega \tag{62}$$

Here $\epsilon^{j\omega t}$ is the complex formulation of the sum of a cosine and sine in quadrature

$$\epsilon^{j\omega t} = \cos \omega t + j \sin \omega t \tag{63}$$

The integration of Eq. (62) is carried out from minus infinite frequency to plus infinite frequency. If the spectrum of the pulse $S(\omega)$ covers only a finite region, then the integration takes place over that region.

Since Eq. (62) has a complex integrand, its application may prove difficult to those whose experience with such matters is not extensive. However, matters are considerably simplified when the transient pulse has a symmetrical shape and the axis of symmetry is placed on the zero time axis (as an even function). Then the sine term in Eq. (63) drops out. Moreover, the spectrum $S(\omega)$ is then a real quantity, since the phase spectrum is zero. Then the integral of Eq. (62) is precisely the Fourier series of Eq. (50), with differential, instead of finite, spacing between the harmonic components and with the coefficients b_k all zero.

FIG. 69.—Transient (nonrepetitive) rectangular pulse, arranged as an even function

To find the spectrum of the nonrepetitive pulse, we use the Fourier integral transform

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} W(t)\epsilon^{-j\omega t}\, dt \tag{64}$$

This equation is the integral formulation of Eqs. (52) and (53), which give a_k and b_k in the repetitive case. Here again the use of an even function, where possible, simplifies the complex formulation by forming a real integrand.

As an example of this process, let us consider the nonrepetitive rectangular pulse, shown in Fig. 69. As in the previously considered repetitive case, we take the maximum amplitude as A volts or amp and the pulse width as d sec. The waveform is an even function, and therefore the sine terms in Eq. (63) are

zero. The limits of the integration are $+d/2$ and $-d/2$. Substituting these values in Eq. (64), we obtain

$$S(\omega) = \frac{A}{2\pi} \int_{-d/2}^{+d/2} \cos \omega t \, dt \tag{65}$$

This integral is

$$S(\omega) = \frac{Ad}{2\pi} \frac{\sin \omega d/2}{\omega d/2} \tag{66}$$

This expression is the complex formulation of the spectrum of the nonrepetitive pulse. Since $S(\omega)$ in this case is a real quantity, it follows that the phase spectrum is zero. Hence Eq. (66)

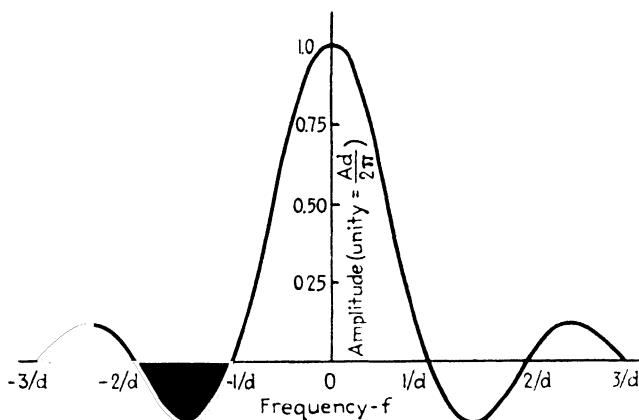



FIG. 70.—Amplitude spectrum of the transient rectangular pulse, a continuous distribution of differentially adjacent harmonics. Compare with Fig. 67.

expresses the envelope of the continuous amplitude spectrum. Here again, we note that the spectrum has the form $\sin x/x$. In fact this part of the equation is identical to that governing the repetitive case, Eq. (59), with $kf = \omega/2\pi$ substituted. But the amplitude factor in Eq. (66) $Ad/2\pi$ is quite different from that of the repetitive case. The amplitude of the fundamental component (the frequency differentially adjacent to zero frequency) is the area of the pulse Ad divided by 2π.

The amplitude spectrum given by Eq. (66) is plotted in Fig. 70 for negative as well as positive values of the frequency $f = \omega/2\pi$. The zeroes of the pulse spectrum, as in the repetitive case, occur at frequencies that are multiples of the inverse pulse length, that is, $1/d$, $2/d$, $3/d$, etc.

The negative frequencies have no physical significance when

the spectrum of a pulse is considered. Mathematically the negative frequencies are the conjugates of the positive frequencies, and the amplitude of a given frequency is the sum of the two values at the negative and positive values of that frequency.

When the transient pulse is not an even or odd function, the spectrum $S(\omega)$ is complex, and both terms of Eq. (63) must be included. The integration is performed as the vector sum of two quantities, one real and the other imaginary. Fortunately this is a matter of academic interest only, since the general properties of pulses and circuits can generally be deduced through the study of symmetrical, and hence even-function, waveforms.

48. Computation of System Functions. Single-terminal-pair Networks.—Having established the basic methods of determining the spectra of pulse waveforms, the next step is the determination of the amplitude and phase system functions of pulse-transmitting networks. Networks fall naturally into two categories of difficulty in this respect. The simpler by far is the network whose input and output terminals are common, that is, the so-called "two-terminal" network or more accurately the "one-terminal-pair" network. A typical example is the coupling connection between video amplifiers at high frequencies. This coupling connection has but two terminals, since the coupling capacitor is effectively a short circuit at high frequencies. Thus the plate of one stage and the grid of the next are common as one terminal, and ground is the other terminal.

In the one-terminal-pair network, the system function is the impedance or admittance appearing across the terminals. The impedance or admittance is formed by combining all the circuit elements present in the network. The capacitive and inductive elements contribute an amount of reactance that depends on the frequency. This factor introduces changes in the amplitude and phase system functions, as functions of frequency. A purely nonreactive network, on the other hand, displays no amplitude or time-delay discrimination throughout the range of frequencies over which it may be considered nonreactive. Unfortunately the range of frequency, forced to our attention by the short pulses of radar work, extends so high that even a simple carbon resistor and its leads display appreciable capacitive reactance in high impedance circuits.

To find the system function of a one-terminal-pair network, we recall the rules for adding circuit elements in series and parallel, as well as the resistance and reactance expressions of the basic circuit elements. These are briefly

1. To obtain the impedance of circuit elements in series, add resistance terms together and reactance terms together,

$$Z_s = R_1 + R_2 + R_3 + \cdots + j(X_1 + X_2 + X_3 + \cdots) \quad (67)$$

2. To obtain the impedance of circuit elements in parallel add conductances together and susceptances together. The result is the reciprocal of the impedance

$$Y_p = \frac{1}{Z_p} = G_1 + G_2 + G_3 + \cdots + j(B_1 + B_2 + B_3 + \cdots) \quad (68)$$

3. Resistance has an impedance equal to its resistance value and admittance equal to its conductance value

$$Z_r = R \quad (69)$$

$$Y_r = G = \frac{1}{R} \quad (70)$$

4. Inductance has a reactance equal to the inductance times the angular frequency and susceptance equal to the inverse of reactance.

$$X_l = \omega L \quad (71)$$

$$B_l = \frac{1}{\omega L} \quad (72)$$

5. Capacitance has reactance equal to the inverse of the product of capacitance and angular frequency, and susceptance equal to the inverse of reactance

$$X_c = \frac{1}{\omega C} \quad (73)$$

$$B_c = \omega C \quad (74)$$

6. When a network consists of series and parallel groups, the parallel elements are reduced to series equivalents that are added to adjacent series elements.

When the impedance or admittance of the network between the terminals has been thus determined, it is expressed in the conventional "$a + jb$" form.

$$Z = R_e + jX_e \quad (75)$$

or

$$Y = G_e + jB_e = \frac{R_e - jX_e}{R_e^2 + X_e^2} \tag{76}$$

Where R_e, G_e, X_e, and B_e are, respectively, the equivalent series resistance, shunt conductance, series reactance, and shunt susceptance of the network. The amplitude system function is then given as

$$|Z_e| = \sqrt{R_e^2 + X_e^2} \tag{77}$$

or

$$|Y_e| = \sqrt{G_e^2 + B_e^2} \tag{78}$$

and the phase system functions are

$$\theta_Z = \tan^{-1} \frac{X_e}{R_e} \tag{79}$$

or

$$\theta_Y = \cot^{-1} \frac{B_e}{G_e} \tag{80}$$

It will be noted that the system functions have two forms, one in impedance terms, the other in admittance. Which form is used depends on the method of exciting the network. If a voltage pulse is applied to the network, the voltage pulse spectrum is combined with the admittance system functions to obtain the resulting current in the network ($I = YE$). If a current pulse is applied, the impedance system function is used to determine the resulting voltage drop ($E = ZI$).

These rules apply generally to any conceivable circuit made up of R, L, and C elements, with any two points in the network selected as the terminals. The rules are simple, but the mathematical forms of the impedance or admittance functions are forbiddingly complex, even in simple circuits.

49. System Functions of the Video Coupling Connection.—We have already treated the simplest two-terminal networks by the transient method, with the results shown in Fig. 51. We now consider a circuit almost as simple as these, to illustrate the steady-state method. The circuit is the shunt-tuned circuit with resistance in the inductive branch, shown in Fig. 71. This circuit is the equivalent plate load circuit, at high frequencies, of the shunt-compensated video amplifier. It is thus often

used in pulse transmitting systems. In what follows we shall adopt the impedance terminology, since the preceding video amplifier tube has a high internal resistance and may be considered as a current source. The output voltage is the current pulse applied times the impedance of the circuit shown.

FIG. 71.—Prototype of the high-frequency video-amplifier coupling connection, a typical one-terminal-pair network.

The resistive-inductive branch of Fig. 71 is first combined to form the equivalent branch impedance

$$Z_1 = R + j\omega L \qquad (81)$$

The other branch impedance is

$$Z_2 = \frac{1}{j\omega C} \qquad (82)$$

The resultant impedance of the two branches, Z_e, is

$$Z_e = \frac{Z_1Z_2}{Z_1 + Z_2} = \frac{(R + j\omega L)(1/j\omega C)}{R + j(\omega L - 1/\omega C)} \qquad (83)$$

$$= \frac{R + j\omega L}{1 - \omega^2LC + j\omega CR} \qquad (84)$$

Equation (84) is rationalized by multiplying numerator and denominator by $1 - \omega^2LC - j\omega CR$. The result is

$$Z_e = \frac{R + j(\omega L - \omega^3L^2C - \omega CR^2)}{(1 - \omega^2LC)^2 + (\omega CR)^2} \qquad (85)$$

The magnitude of this complex impedance (the amplitude system function) is

$$|Z_e| = \frac{\sqrt{R^2 + (\omega L - \omega^3L^2C - \omega CR^2)^2}}{(1 - \omega^2LC)^2 + (\omega CR)^2} \qquad (86)$$

and the phase system function is

$$\theta_z = \tan^{-1} \frac{(\omega L - \omega^3L^2C - \omega CR^2)}{R} \qquad (87)$$

As promised, these expressions are complicated, although the circuit is simple. By means of Eqs. (86) and (87) we can plot

the amplitude and phase system functions throughout all the required values of the angular frequency, provided that we assign particular numerical values of L, C, and R to the circuit. But such numerical examples do not take into account the necessary properties of the circuit for its intended purpose, coupling video amplifier stages at high frequencies. It is necessary, therefore, to establish relations between L, C, and R that will best serve the intended purpose. The system functions can then be recast in general form and plotted as functions of frequency.

The clue to the desired characteristics is the statement, previously made, that a video amplifier must display constant amplification and constant time delay over the range of significant frequencies occupied by the pulse spectrum. It follows that the L,C,R circuit of Fig. 71 must display a constant impedance from the lowest frequencies to the highest.

The shunt capacitance reduces the impedance as the frequency exceeds the resonant frequency of the circuit. Hence the resonant frequency must be kept at least as high as the maximum frequency in the pulse spectrum. This implies the use of the smallest possible value of C and of a similarly small value of L.

No restrictions have been placed, thus far, on the value of R. We recall also that the impedance must be the same at the lowest frequencies as at the highest. Since L and C must be small to uphold the impedance value at high frequency, it follows that they have negligible effect at the lowest frequencies. In other words, the impedance of the circuit at a very low frequency is simply R ohms. This value of impedance must be maintained, then, up to the highest significant value of frequency in the pulse spectrum.

The relationships among L, C, and R may be further identified by noting that the impedance of the resistive-inductive branch increases with frequency while that of the capacitive branch decreases. The two effects can be made to balance, approximately, if the susceptances of the two branches are made equal.

$$B_{RL} = B_C \tag{88}$$

$$\frac{\omega L}{R^2 + (\omega L)^2} = \omega C \tag{89}$$

Since L is very small we may neglect $(\omega L)^2$ compared with R^2.

Then Eq. (89) reduces to

$$L = CR^2 \tag{90}$$

This is the approximate condition for balancing the reactive effect of one branch against that of the other.

We require one additional relationship between L, C, and R to obtain a general statement. We select as a reference angular frequency ω_0, the frequency at which the resistance equals the capacitive reactance

$$R = \frac{1}{\omega_0 C} \tag{91}$$

or

$$\omega_0 = \frac{1}{CR} \tag{92}$$

This is also the resonant frequency of the circuit when L is related to R and C by Eq. (90). We may expect therefore that the impedance will be maintained constant, within reasonable limits, up to the frequency ω_0. The reference frequency is chosen equal to the highest significant frequency in the pulse spectrum. When ω_0 has been set by this consideration and C is the smallest capacitance possible (wiring and tube capacitance only), then a maximum value is placed on R by Eq. (91).

We now have the basic considerations for the choice of L, R, and C in this circuit. The capacitance must have the smallest possible value. The highest frequency ω_0 in the pulse spectrum to be transmitted without discrimination is then specified, and from ω_0 and C, R is determined by Eq. (91). Then, with R and C fixed, L is determined from Eq. (90) or from some equivalent expression [values of L lower than those given by Eq. (90) are used in practice, as we shall see later].

The foregoing is an elementary problem in network synthesis. Having chosen L and R in terms of C and ω_0, it remains to determine the network system functions. When Eqs. (90) and (91) are substituted in the system function Eqs. (86) and (87), relations result that can be plotted in general terms. The substitution results in the following system functions:

$$|Z_e| = \frac{R(1 - j(\omega^3/\omega_0^3))}{1 - \omega^2/\omega_0^2 + \omega^4/\omega_0^4} \tag{93}$$

and

$$\theta_Z = \tan^{-1} - \frac{\omega^3}{\omega_0^3} \tag{94}$$

We note that the ratio of the particular frequency considered to the reference frequency, ω/ω_0, appears frequently in these

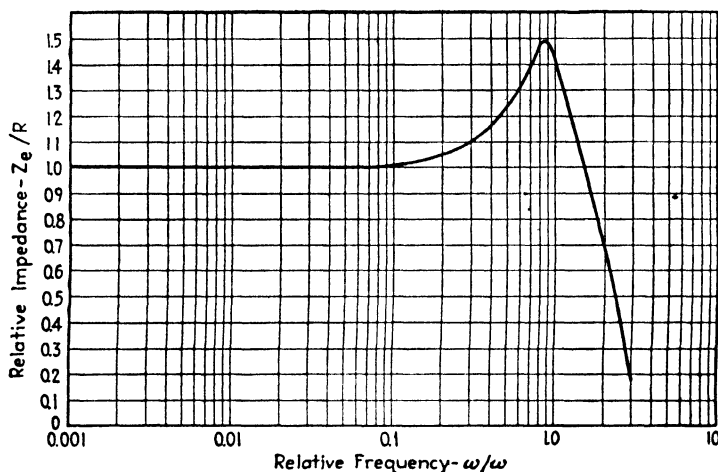



FIG. 72.—Amplitude system function of the circuit in Fig. 71, when L is chosen to equal CR^2.

equations, in the second, third, and fourth powers. It is thus appropriate to plot the system functions in terms of relative frequency ω/ω_0. Such plots appear in Figs. 72 and 73.

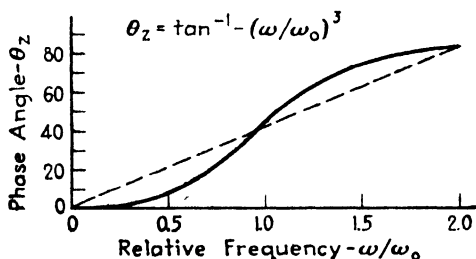



FIG. 73.—Phase system function corresponding to Fig. 72.

It is evident that our choice of L, R, and C is imperfect since the system functions based on this choice are far from ideal. The resonance peak of Fig. 72 is too prominent, and the phase system function is far from linear. To reduce the resonance peak and smoothen the phase curve, it is necessary to reduce the ratio

$Q = \omega_0 L/R$. The quantity Q is the well-known circuit-merit figure discussed at length in Chap. III, where it is shown that reduction of Q produces a flattening of the resonance curve of a

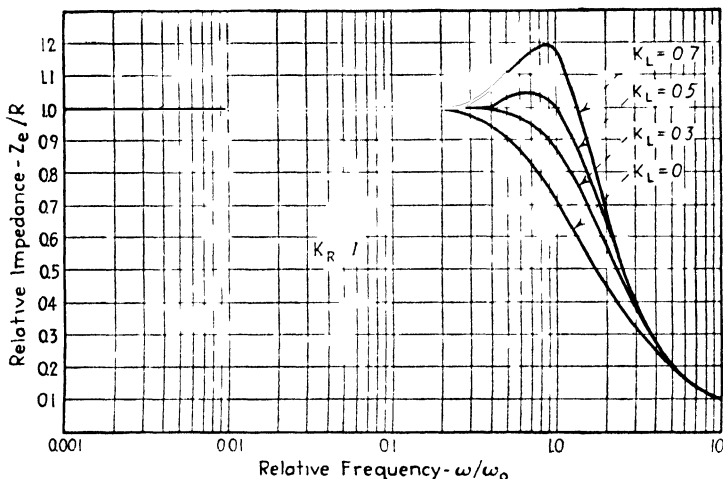



FIG. 74.—Amplitude system functions of circuit in Fig. 71, when L equals $K_L C R^2$

tuned circuit. The Q may be reduced by decreasing the value of L or by increasing the value of R. Increasing the value of R has the additional effect of increasing the impedance Z_e at all frequencies, which may be desirable, particularly where high gain is required in the video amplifier.

To take into account variations in the values of L or R we rewrite Eqs. (90) and (91) with factors K_L and K_R, respectively

$$L = K_L C R^2 \qquad (95)$$

and

$$R = K_R \frac{1}{\omega_0 C} \qquad (96)$$

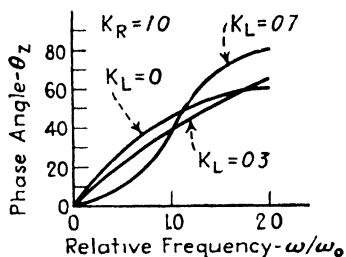



FIG. 75.—Phase system functions corresponding to Fig. 74.

When these expressions are substituted in Eq. (92) the following rather formidable amplitude system function results:

$$|Z_e| = \frac{R(1 + j(K_L/K_R - K_R)(\omega/\omega_0) - j(K_L^2/K_R)(\omega^3/\omega_0^3))}{1 - (2K_L - K_R^2)(\omega^2/\omega_0^2) + K_L^2(\omega^4/\omega_0^4)} \qquad (97)$$

The phase system function is

$$\theta_Z = \tan^{-1}\left(K_L/K_R - K_R\frac{\omega}{\omega_0} - \frac{K_L^2}{K_R}\frac{\omega^3}{\omega_0^3}\right) \tag{98}$$

Figures 74 through 77 show the amplitude and phase functions

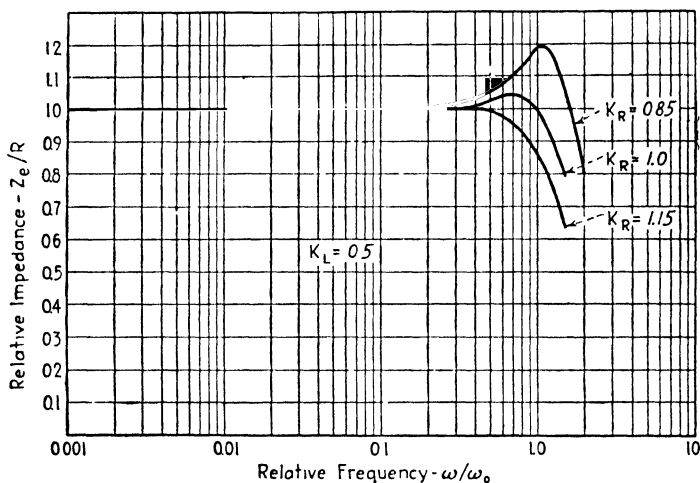



FIG. 76.—Amplitude system functions of circuit in Fig. 71, when R equals $K_R/(\omega_0 C)$.

of this circuit, as functions of ω/ω_0 for various values of K_L and K_R. It will be noted that decreasing K_L (lowering the inductance value) lowers the resonance peak and smoothens the phase curve. Similar effects occur as K_R is increased (increasing the resistance value). Typical values in practice are $K_R = 1.0$ and $K_L = 0.5$ for amplifiers containing few stages. The values $K_R = 0.85$ and $K_L = 0.3$ produce particularly linear results and are often used in multistage amplifiers. The value $K_R = 0.85$ implies a 15 per cent loss in stage gain relative to the amplifier with $K_R = 1$.

FIG. 77.—Phase system functions corresponding to Fig. 76.

50. System Functions of a Two-terminal-pair Network.—When the network has separate input and output terminals it is called

a "two-terminal-pair network" (commonly, a "four-terminal" network). The computation of the system functions in this case is considerably more complicated, so much so that computation is usually dispensed with in favor of measurements. Here we shall indicate only the general approach.

The two-terminal-pair network shown in Fig. 78 is a box within which exist circuit elements represented symbolically by the admittance Y_{12}. The voltage source contains an internal impedance Z_s. It is connected across the input terminals and develops across them a sinusoidal voltage of amplitude E_i.

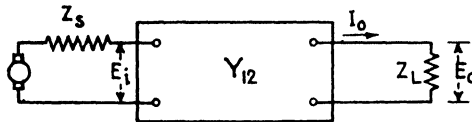



FIG. 78.—Basic two-terminal-pair network and external impedances.

The transfer admittance is so defined that the corresponding sinusoidal current I_0 appearing at the output terminals is

$$I_0 = E_i Y_{12} \tag{99}$$

The output voltage is the product of the output current times the load impedance Z_L.

$$E_0 = I_0 Z_L = E_i Y_{12} Z_L \tag{100}$$

Hence the voltage gain (or loss) of the network is

$$\frac{E_0}{E_i} = Y_{12} Z_L \tag{101}$$

Similarly a transfer impedance Z_{12} may be defined such that the ratio of output to input current (the current gain or loss) is given by

$$\frac{I_0}{I_i} = \frac{Z_{12}}{Z_L} \tag{102}$$

The transfer admittance or impedance and load impedance are evidently the system functions that describe the network. The amplitude of Eq. (101) or (102) is the amplitude system function and its phase angle is the phase function. Both functions are generally complex expressions involving the L, C, and R components and the frequency.

When Y_{12} or Z_{12} and Z_L are computed or measured as func-

tions of frequency, it is possible to determine the response of the network to an applied pulse by treating each significant frequency in the pulse spectrum separately and by adding the sinusoidal responses. Thus if a_k is the voltage amplitude of the kth harmonic in the pulse spectrum and the pulse phase spectrum is zero (an even-function repetition pulse), the output voltage of the network is

$$e_0(t) = \sum_{k=0}^{k=\infty} a_k Y_{12} Z_L \cos k2\pi ft \tag{103}$$

If the pulse spectrum is continuous (transient pulse), the equivalent expression is

$$e_0(t) = \int_{-\infty}^{\infty} S(\omega) Y_{12} Z_L e^{j\omega t} \, d\omega \tag{104}$$

where $S(\omega)$ is the continuous spectrum of the applied pulse and $Y_{12}Z_L$ is the system function in terms of the frequency ω.

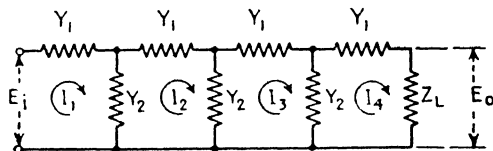



FIG. 79.—Two-terminal-pair network composed of like shunt and like series elements (typical filter structure).

Among the important two-terminal-pair networks is the wave filter. Filter design is greatly simplified by the use of similar impedances in the shunt and series branches. A typical ladder type filter is shown in Fig. 79. The elements are shown as admittances, all the series elements being identical and all the shunt elements being identical. A sinusoidal voltage E_i is applied at the input and a sinusoidal current I_4 flows in the load impedance. In each mesh a current flows in accordance with the net voltage applied from one mesh to the next.

The four mesh current are written in four simultaneous equations, as follows:

$$E_i = \frac{I_1}{Y_1} + \frac{I_1 - I_2}{Y_2} \tag{105}$$

$$\frac{I_1 - I_2}{Y_2} + \frac{I_2}{Y_1} + \frac{I_2 - I_3}{Y_2} = 0 \tag{106}$$

$$\frac{I_2 - I_3}{Y_2} + \frac{I_3}{Y_1} + \frac{I_3 - I_4}{Y_2} = 0 \qquad (107)$$

$$\frac{I_3 - I_4}{Y_2} + I_4\left(Z_L + \frac{1}{Y_1}\right) = 0 \qquad (108)$$

When these equations are solved and I_4 is found as a function of frequency, the transfer admittance is given by

$$Y_{14} = \frac{I_4}{E_1} \qquad (109)$$

When Y_1 and Y_2 are single capacitors and inductors ("constant-k" filter), the solution of the equations is not difficult. But when the elements are more complex combinations of inductance and capacitance, and if the quantity of each in each mesh differs, it is usually impossible to deduce any general properties of the network from the resulting expressions. It is then necessary to resort to plotting the system functions for particular values of the L, C, and R elements.

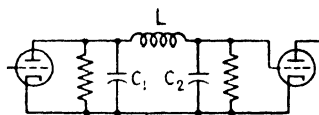



FIG. 80.—Series-compensated video-amplifier coupling connection, a typical two-terminal-pair network.

An important type of two-terminal-pair network, related to the preceding one-terminal-pair example, is the "series-compensated" video amplifier coupling connection shown in Fig. 80. Here the series inductive element separates the input and output terminals and also separates the capacitances at the two ends of the circuit. This permits better h-f performance than that of the simple "shunt-compensated" circuit shown in Fig. 71. The properties of these and other video amplifier circuits are discussed in Chap. VI.

51. System Functions of the Ideal Low-pass Filter.—As an example of the two-terminal-pair network, we shall choose, as most generally informative, the so-called "ideal" low-pass filter. This filter has the amplitude and phase system functions shown in Figs. 81 and 82. It transmits without amplitude discrimination all frequencies from zero to an upper limit ω_0, and it introduces a phase shift directly proportional to frequency over the same limits. As a mathematical convenience the system functions are shown symmetrically about zero frequency, extend-

ing from $-\omega_0$ to $+\omega_0$. This has no effect on the result, since we are concerned only with positive frequencies.

The system function of this filter is described as

$$Y_{12}Z_L = K\epsilon^{-j\omega t_0} \tag{110}$$

where K is the constant amplitude response of Fig. 81 and t_0 is the slope of the linear phase characteristic of Fig. 82.

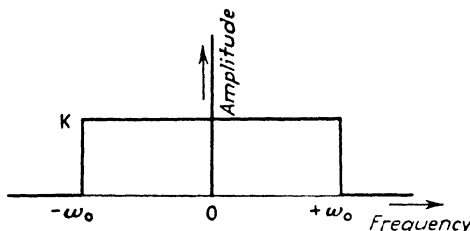



FIG. 81.—Amplitude system function of the ideal low-pass filter.

Substituting Eq. (110) in (104) we write

$$e_0(t) = \int_{-\infty}^{\infty} S(\omega)K\epsilon^{-j\omega t_0}\epsilon^{j\omega t}\, d\omega \tag{111}$$

$$= K \int_{-\infty}^{\infty} S(\omega)\epsilon^{j\omega(t-t_0)}\, d\omega \tag{112}$$

But, by comparison with Eq. (62), this integral is equal to the applied waveform $e_s(t)$, multiplied by the constant K and dis-

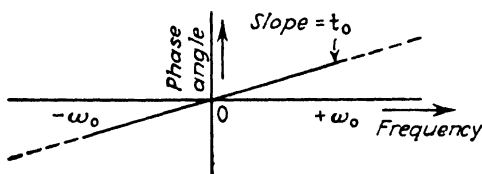



FIG. 82.—Phase system function of the ideal low-pass filter.

placed in time by an amount t_0. Thus

$$e_0(t) = Ke_s(t - t_0) \tag{113}$$

This remarkable result indicates that the ideal low-pass filter passes any applied waveform without distortion in shape, merely changing its size and delaying it in time. The delay is equal to the slope of the phase system characteristic. This result occurs so long as the maximum pass frequency of the filter ω_0 is

higher than the highest significant frequency in the pulse spectrum.

The reason for calling this filter "ideal" is now evident: such a filter introduces no change in the shape of the transmitted pulse (or any waveform). The criteria of such distortionless reproduction are simple: no amplitude discrimination throughout the spectrum of the applied waveform and linear phase shift in the same region.

52. Steady-state Computation of Output Waveforms.—We now come to the final step in the steady-state analysis of networks, the determination of the size, shape, and position in time of the output waveform. The process, stated explicitly in Eqs. (103) and (104), is evident from the preceding discussion: the sinusoidal components of the applied waveform (its amplitude and phase spectra) are combined with the system functions. Multiplication of the input amplitude spectrum by the amplitude system function gives the amplitude spectrum of the output pulse. Addition of the input phase spectrum to the phase system function gives the phase spectrum of the output pulse. Addition of the output spectral components, in proper relative amplitudes and phases, generates the output waveform in its proper size, shape, and position in time.

The general statement of this process is given in the integral

$$W_0(t) = \int_{-\infty}^{\infty} N(\omega)S(\omega)\epsilon^{j\omega t}\,d\omega \qquad (114)$$

Here $N(\omega)$ is the system function of the network, a complex quantity having an amplitude and a phase. $N(\omega)$ is the self-admittance or self-impedance of a single-terminal-pair network; it is the product of the transfer admittance and load impedance (or quotient of the transfer impedance by the load impedance) of a two-terminal-pair network. $S(\omega)$ is the complex spectrum of the applied pulse. The quantity $\epsilon^{j\omega t}$ is the sum of a sine and cosine term of frequency ω in quadrature. The sum of these sinusoidal components, with the indicated amplitudes and phases, over an infinite frequency range (less than infinite when the pulse spectrum is less than infinite) gives the output waveform $W_0(t)$ as a function of time.

We shall examine two examples of this process: one in which the pulse spectrum comprises discrete harmonics, the other continuous.

53. Pulse Transmission by the Video Amplifier Coupling Connection.—As the first example consider a rectangular pulse (even function) applied to the compensated video amplifier connection of Fig. 71. We shall consider two cases: one in which the transmission system displays sharp cutoff ($K_L = 1$, $K_R = 1$, Fig. 72), the other more gradual cutoff ($K_L = 0$, $K_R = 1$, Fig. 74). These amplitude system functions are multiplied by the rectangular pulse amplitude spectrum, with the result shown in Fig. 83. We assign the pulse width as $d = 1/\omega_0$, so that its

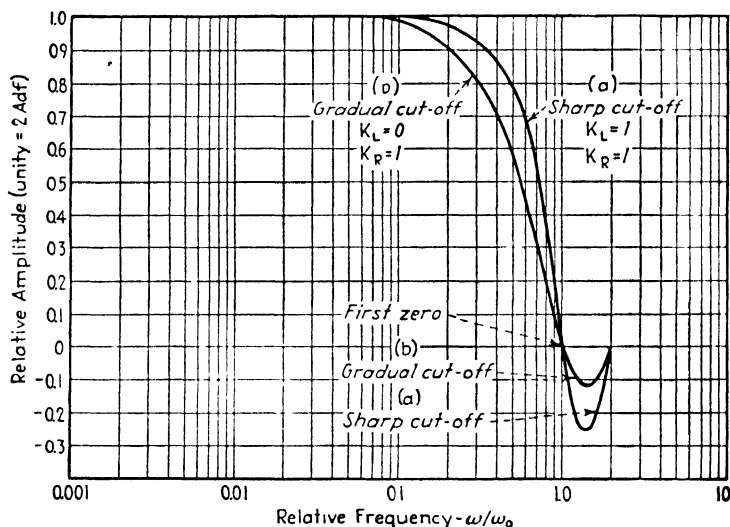



FIG. 83.—Amplitude spectra of rectangular pulse after transmission through network of Fig. 71, adjusted for sharp cutoff and gradual cutoff conditions.

first zero occurs at a relative frequency $\omega/\omega_0 = 1$. The phase system functions for the two cases are obtained from Figs. 73 and 75. These are added to the pulse phase spectrum. Since this is $\theta = 0$, the phase spectra of the reproduced pulse are given directly in Figs. 73 and 75.

To obtain the output waveform from these data, we employ Eq. (103) and add the first 10 harmonics with the amplitudes given by Fig. 83 and the phases given by Figs. 73 and 75. The resulting waveforms are shown in Figs. 84 and 85. The pulses shown differ from those actually reproduced by the circuit since harmonics above the tenth have been neglected, but the approximation is close.

These waveforms show typical phenomena associated with pulse transmission. (1) In both cases, the pulse has lost its rectangular shape, and the sides of the pulse rise and fall gradually. (2) The pulse is broadened at the base relative to the width of the applied pulse. (3) The pulse is delayed slightly in time.

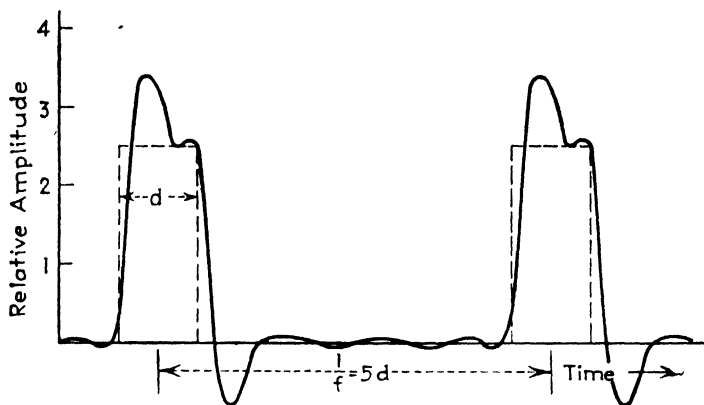



FIG. 84.—Rectangular pulse after transmission through a sharp cutoff filter, synthesized from ten harmonics corresponding to *a* in Fig. 83.

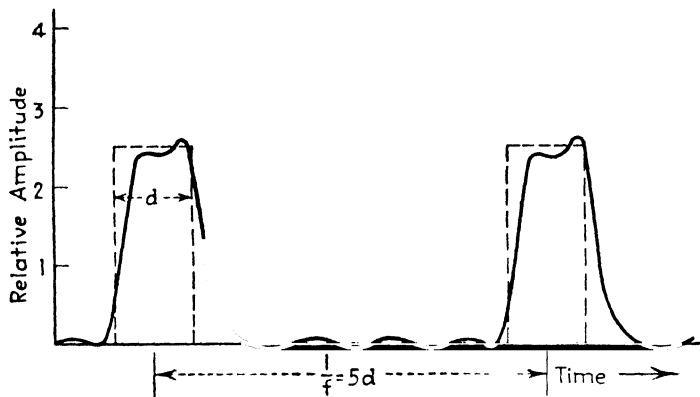



FIG. 85.—Rectangular pulse after transmission through a gradual cutoff filter, synthesized from ten harmonics corresponding to *b* in Fig. 83.

A most important difference between the two cases is the shape of the pulse at the top of the leading edge and the bottom of the trailing edge. In Fig. 84, representing the sharp-cutoff case, the pulse "overshoots" at top and bottom due to the higher energy storage associated with the sharp cutoff. In the gradual cutoff case (Fig. 85) this overshoot is absent. But, also, in Fig.

85 the leading and trailing edges of the pulse rise and fall more slowly than in the sharp-cutoff case. These two effects are generally associated. Sharp cutoff produces a steeply rising and falling pulse at the expense of overshoot. More gradual cutoff removes the overshoot, but, for a given value of ω_0, responds more slowly to the rise and fall of the pulse.

54. Pulse Transmission by the Ideal Low-pass Filter.—As a second example of pulse transmission, we shall take a case in which analytic expression is possible. We return to the ideal low-pass filter of Figs. 81 and 82. We shall apply the unit step voltage of Fig. 40*a* to this filter and compute the response. To do so we employ Eq. (114). The spectrum $S(\omega)$ of the unit step voltage is the time integral given by Eq. (64). Since the unit voltage extends to infinite time, its integral with respect to time must be found by a mathematical subterfuge.[1] The result is

$$S(\omega) = \lim_{a \to 0} \frac{1}{2\pi(a + j\omega)} \tag{115}$$

The system function of the ideal low-pass network is Eq. (110). Substituting Eqs. (115) and (110) in Eq. (62), we have

$$e_0(t) = \lim_{a \to 0} \int_{-\omega_0}^{+\omega_0} \frac{K\epsilon^{-j\omega t_0}\epsilon^{j\omega t}}{2\pi(a + j\omega)}\, d\omega \tag{117}$$

The complex exponential term $\epsilon^{j\omega(t-t_0)}$ is expanded as the sum of a real cosine term and an imaginary sine term of argument $\omega(t - t_0)$. The integral of the cosine terms, in the limit as a approaches zero, is $K/2$. The integral can then be written

$$e_0(t) = \frac{K}{2} + \frac{K}{2\pi} \int_{-\omega_0}^{\omega_0} \frac{\sin \omega(t - t_0)}{\omega}\, d\omega \tag{118}$$

By a change of variables $u = \omega(t - t_0)$, this can be converted to

$$e_0(t) = \frac{K}{2} + \frac{K}{2\pi} \int_{-u}^{u} \frac{\sin u}{u}\, du \tag{119}$$

[1] The spectrum of the unit step voltage is found by taking the unit voltage as ϵ^{-at} for all values of t greater than $t = 0$, and allowing a to approach zero. Thus, by Eq. (64)

$$S(\omega) = \frac{1}{2\pi} \int_0^\infty \epsilon^{-at}\epsilon^{-j\omega t}\, dt = \frac{1}{2\pi(a + j\omega)} \tag{116}$$

This result is correct only in the limit as a approaches zero, as stated in Eq. (115).

We thus find that the output waveform is proportional to the area under the $(\sin u)/u$ curve (Fig. 70), plus a constant. The integral of $(\sin u)/u$, abbreviated $Si(u)$, is available in tables. From the tabulated values the waveform can be plotted, noting that $t = (u + \omega t_0)/\omega$, as shown in Fig. 86. This is the response of the ideal low-pass filter to the unit step voltage.

The waveform of Fig. 86 is of basic interest in the study of pulse transmission systems because it shows clearly the relation between the changes of pulse shape introduced by the network and the characteristics of the network. First it will be noted that the pulse is delayed, at the half-voltage level, by a time t_0. Thus the slope of the system phase function gives directly the time delay introduced by the network.

FIG. 86.—Reproduction of a unit step function after passage through the ideal low-pass filter (Figs. 81 and 82).

The time of rise of the leading edge of the pulse, indicated by the limits of the dashed lines, is t_r sec. This time represents the degree of "sluggishness" of the network in responding to the infinitely-sharp leading edge of the applied pulse. The value of t_r is found by differentiating Eq. (119) with respect to time.

$$\frac{de_0}{dt} = \frac{K\omega_0}{\pi} \tag{120}$$

In rising to the amplitude K the time required is

$$t_r = \frac{\pi}{\omega_0} = \frac{1}{2f_0} \tag{121}$$

Thus the time of rise of the pulse is equal to one half the inverse of the cutoff frequency of the filter

This result shows that the minimum band of frequencies required for transmitting a pulse of width $d = 2t_r$ is the inverse of the pulse length. Thus by Eq. (121), the required upper frequency limit is

$$f_0 = \frac{1}{2t_r} = \frac{1}{d} \tag{122}$$

It should be noted that when such a bandwidth is employed, the pulse has the slowly rising leading and trailing edges indicated

by Fig. 86. A wider band is required to match the output pulse edges more closely to the applied pulse. This is indicated in Fig. 87, in which the time scale has been compressed (and the frequency scale correspondingly widened). It is thus evident that the fidelity of pulse transmission improves as the bandwidth is widened.

The oscillations at the base and top of the pulse occur as the integration [Eq. (119)] is carried through the zeroes of the $(\sin u)/u$ curve. Since the zeroes are separated at intervals of $2\pi/\omega_0$, it follows that the period of the oscillations has the same value. These oscillations are closely related to the overshoot oscillations encountered in Fig. 84, formed by summing

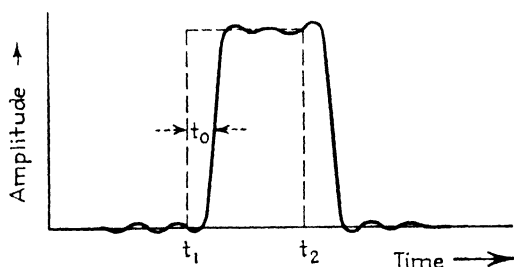



FIG. 87.—Reproduction of rectangular pulse by ideal low-pass filter.

up the harmonics of a repetitive pulse. Their amplitude may be reduced, as in Fig. 85, by the use of a system amplitude function having a more gradual cutoff than that of the ideal filter here considered.

One final aspect of Fig. 86 must be considered. It will be noted that the base oscillations extend to the left past time $t = 0$ and into negative values of time. The implication is that the ideal low-pass filter begins to respond to a pulse before it is applied! Since this is a physically impossible result, we conclude that the ideal system functions (Figs. 81 and 82) of the filter cannot be realized in practice. This is indeed the case. When the system functions of a realizable low-pass filter are considered, it is found that the integration of these functions produces no response before time $t = 0$, as is required by physical reasoning. However, certain practical filters do produce a response prior to the main output pulse rise. These base oscillations are akin to those shown in Fig. 86, and arise from the same considerations. It

follows that the time delay of the filter t_0 applies to the main pulse response but not to the whole response of the circuit.

55. Relative Effects of Amplitude and Phase Distortion.—The system functions of the low-pass filter just considered are ideal in that the filter will reproduce without change of shape any pulse whose spectrum is wholly contained within the frequency limits zero and ω_0. If the pulse spectrum exceeds the upper limit, the filter action is no longer ideal, and the pulse shape is distorted by the transmission process. Such distortion is considered in two categories: amplitude distortion and phase

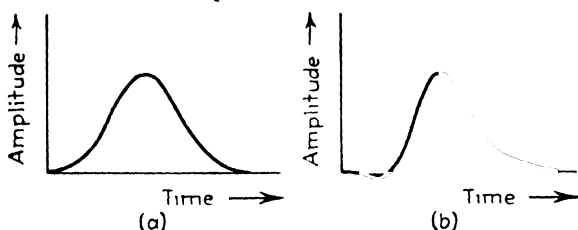



FIG. 88.—Symmetrical (a) and unsymmetrical (b) distortion of pulse reproduction due to amplitude distortion only and phase distortion only, respectively.

distortion. Amplitude distortion arises in a network whose amplitude system function is not uniform throughout the pulse spectrum. Phase distortion arises when the phase system function is not linear over the same region. In general, when both types of distortion are present, it is very difficult to identify them by inspection of the output pulse.

Theoretical considerations show, however, that two separate and distinct effects are caused by the presence of either type of distortion in the absence of the other. If only amplitude distortion is present, and the phase system function is ideal, the distortion of the pulse shape is symmetrical about its center line. On the other hand, if the phase distortion is present, and the amplitude system function is ideal, the distortion occurs asymmetrically about the center line of the pulse. The two cases are shown in Fig. 88.

Since amplitude distortion produces an output pulse that is an even function of time, it may be analyzed as a sum of cosine terms. But when phase distortion is present, the pulse is generally neither an even nor an odd function, and the analysis must be carried out in the complex plane. For this reason, little is

known about the general effects of phase distortion except as they may be computed in simple cases or measured in particular transmission systems. It is known that the nonlinear phase characteristic which is present in all transmission systems cannot be ignored, but the only general method of taking it into account is the series method, which yields results only for particular values of the parameters.

The difficulty of phase distortion computation has led to a general disregard of its effects in design, with much resulting confusion. Thus for some years in the early history of radar development, it was believed that the amplitude system function should be flat and uniform over the frequency range covered by the pulse spectrum and that the cutoff outside this limit could be of any convenient shape. Later it was realized that the phase characteristic associated with the amplitude function is excessively nonideal if the amplitude cutoff is sharp. This nonideal phase characteristic produces overshoot oscillations of high amplitude and may, in addition, prolong the output pulse considerably beyond the length of the applied pulse. It then became clear that a nonideal amplitude system function (that is, one not flat over the pulse spectrum) could be tolerated and was in fact desirable, provided that the phase system function was thereby made more closely linear. It is usually true that a gradual cutoff in the amplitude function is associated with a linear phase function in the same region.

When practical use was made of this fact, it was found that the overshoot oscillations could be reduced (and the length of the transmitted pulse reduced in certain instances) at but slight loss in the steepness of the output pulse edges. We have already noted this general effect in Figs. 84 and 85. One result of this discovery has been the gradual elimination of transformer-coupled i-f amplifier stages, with their steep-sided amplitude response curves, in favor of single-tuned circuits, with more nearly ideal phase characteristics.

56. Transmission of Pulse-modulated Radio-frequency Carriers.—Thus far our study of pulse-transmission systems has been confined to those operated at video frequencies, that is, frequencies from zero up to the maximum frequency in the pulse spectrum. When the pulse waveform is used to modulate an h-f carrier (by amplitude modulation), we are confronted with

transmission at radio or intermediate frequencies and with transmission bands symmetrical about the carrier frequency. This is evidently a more complicated matter, because the amplitude and phase of the carrier frequency must be considered in addition to the frequencies of the pulse envelope. The complication is not serious, however, if the carrier frequency exceeds the maximum frequency in the pulse spectrum by a substantial amount. This always is true in radar transmission at radio frequencies and to a somewhat less extent when transmission occurs at intermediate frequencies.

To study the pulse-modulated carrier, we consider a carrier of angular frequency ω_c and phase ϕ_c, modulated by a pulse waveform $W(t)$. The input voltage applied to the transmission system is their product.

$$e_i(t) = W(t) \cos (\omega_c t + \phi_c) \tag{123}$$

To find the spectrum of this voltage, we employ Eq. (64). The computation is aided by writing the input voltage as the sum of two exponentials.

$$e_i(t) = W(t) \times \tfrac{1}{2}[\epsilon^{j(\omega_c t+\phi_c)} + \epsilon^{-j(\omega_c t+\phi_c)}] \tag{123a}$$

Then Eq. (64) becomes

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(t)\epsilon^{\pm j(\omega_c \pm \omega_m)t}\epsilon^{\pm j\phi_c}\, dt \tag{124}$$

where ω_m is the frequency of the modulating envelope. Comparison of this equation with Eq. (64) shows that the spectrum of the modulated wave is identical to that of the pulse envelope itself, except that the frequencies involved in the spectrum are now the carrier frequency plus or minus the modulating frequencies. This result is a familiar proposition in communications practice. It states simply that when a waveform modulates a carrier in amplitude, the spectrum of the waveform is translated upward from zero frequency to the carrier frequency, and the spectrum appears symmetrically about the carrier in the form of two sidebands.

The term $\epsilon^{\pm j\phi_c}$ in Eq. (124) represents the carrier phase. So long as the phase remains constant, this term is merely a multiplying constant, representing a sine and cosine of the angle ϕ_c

in quadrature. If the carrier phase changes substantially with time, as it does when phase modulation or frequency modulation is present, the carrier spectrum is not identical to that of the

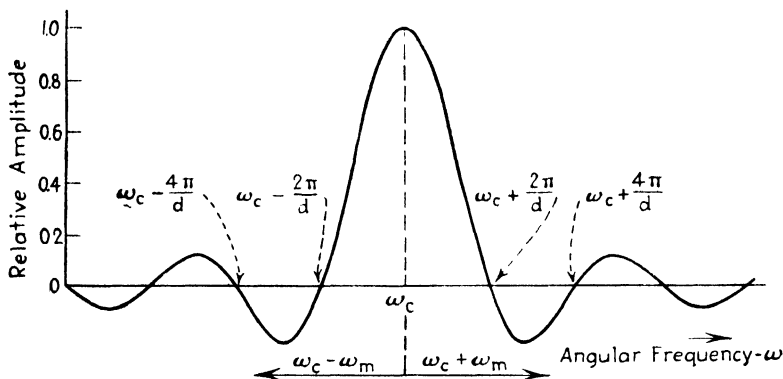



FIG. 89.—Amplitude spectrum of carrier wave modulated with a rectangular pulse (compare with Figs. 67 and 70).

modulating envelope, and no simple relation exists between them.

When the waveform $W(t)$ is a transient rectangular pulse, the spectrum given by Eq. (124) becomes

$$S(\omega) = \frac{Ad}{2\pi} \frac{\sin (\omega_c \pm \omega_m)d/2}{(\omega_c \pm \omega_m)d/2} \tag{125}$$

where A is the amplitude of the envelope, d the pulse width, ω_c the angular carrier frequency, and ω_m the angular sideband frequencies. The spectrum given by this expression is plotted in Fig. 89. The similarity to Fig. 70 is evident.

FIG. 90.—Envelope of carrier wave after passage through ideal bandpass filter (compare with Fig. 86).

When the rectangular pulse modulated wave is applied to a transmission system, Eq. (114) gives the output waveform. In the case of the ideal bandpass filter (system functions given in Figs. 81 and 82), the envelope of the output carrier signal is identical to that of Fig. 86. Figure 90 shows the transmitted carrier

after passage through the ideal filter. It should be noted that in the bandpass case the total bandwidth is $2\omega_0$ compared with ω_0 in the low-pass case. With this change, Eqs. (121) and (122) apply.

57. Noise in Pulse Systems.—Thus far our study of pulse transmission has dealt only with the fidelity of transmission as it relates to the design of the transmission system. We must now consider another aspect of pulse transmission, that is, the *energy level* which must be maintained by the radar signal in order to serve a useful purpose. The required energy level is determined by noise, the random currents and voltages everywhere present in circuits and transmission paths. When the radar signal sinks to a level comparable with that of the noise, the fidelity of transmission is marred. When the signal sinks below the noise level by any substantial amount, the signal becomes lost and cannot be regained.

Noise is a particularly evident evil in radar equipment because the maximum range of a radar is determined by the noise level. The maximum range may be increased by increasing the transmitted power, at substantial cost in weight and power requirements. Or the range may equally well be increased by decreasing the noise level, a more difficult technical problem but one that costs little in the weight and power consumption of the equipment. It is not surprising, therefore, that very great effort has been expended on the study of noise sources and their reduction.

Mention has been made in Chap. I, pages 54–58, of the effect of noise on the maximum range of a radar system. Here we consider noise from the standpoint of the transmission system and go into some detail on the evaluation of noise sources.

58. Bandwidth for Maximum Signal-to-noise Ratio.—The effect of noise is measured in terms of the signal-to-noise power ratio. A general relationship can be derived between the maximum signal-to-noise power ratio and the effective bandwidth of the transmission system. The noise considered is random "thermal" noise, the type generally encountered in uhf and shf radar systems. The impulsive type of noise (ignition interference and static) so commonly encountered on lower frequencies is generally absent in radar.

The power developed by a thermal noise source is uniformly distributed over all frequencies. Hence the total noise power

P_N developed in a transmission system whose amplitude system function is $G(f)$ is

$$P_N = K \int_{f_1}^{f_2} G(f)\, df \tag{126}$$

where K is a constant, $G(f)$ is the gain or amplitude response of the transmission system as a function of the frequency f, and f_1 and f_2 are the limiting frequencies of the transmission system passband. For the moment we shall accept K as a constant, and inquire later (Sec. 60) into its nature and magnitude.

Equation (126) shows that the noise power increases in direct proportion as the band limits are spread apart, the shape of the response curve $G(f)$ remaining unchanged in the process. Thus as we employ wider and wider bandwidths in a radar transmission system, other things being equal, the noise power increases proportionately.

We now inquire what happens to the peak power of the output pulse as the bandwidth is increased. The clue is given in the response of the ideal bandpass filter to a unit step voltage [Fig. 86 and Eq. (121)]. We note that the response reaches full amplitude only after a time equal to one-half the inverse of the bandwidth of the filter. Thus as the bandwidth of the filter is reduced, the time required to reach full amplitude grows longer and eventually the time of rise (Fig. 86) surpasses the duration of the applied pulse. It follows that the response of a narrow-band filter never reaches its maximum when a short pulse is applied to it.

This reasoning indicates that the output voltage or current response of a transmission system to an applied pulse increases, more or less proportionately, as the bandwidth is increased to cover more and more of the pulse spectrum. The output signal power (proportional to the square of the output current or voltage) thus increases as the square of the bandwidth. When this square-law increase in output signal power is compared with the linear increase of output noise power, it is evident that the signal-to-noise power ratio increases as the bandwidth is increased. The moral is clear: to reduce the effect of noise, excessively narrow values of bandwidth are to be avoided.

At the other end of the scale, excessively wide values of bandwidth are likewise to be avoided. This follows from the fact that when sufficient bandwidth is available to allow the filter

to attain full response during the application of the pulse, further widening of the band does nothing to increase the signal power. But further widening of the band continues to increase the noise power, so that the signal-to-noise power ratio decreases as the band is increased above a critical value.

This critical value of bandwidth is evidently a matter of primary importance, since its use emphasizes the peak power (or voltage or current) value of the output pulse relative to the noise. The value of the optimum bandwidth is somewhat less than the inverse of the pulse width $1/d$ for the low-pass case. For the bandpass case, employing double-sideband carrier transmission, the optimum bandwidth is twice this value.

The value of optimum bandwidth has been indicated, rather than proved, by the above argument. The proof lies in integrating Eq. (114) over successively wider and wider frequency limits and noting the trend of the maximum amplitude attained by the waveform. It will be found that the amplitude steadily increases until the integration limits exceed the first zero of the pulse spectrum $S(\omega)$, that is, at frequency $1/d$. As successively high zeroes in the spectrum are included within the integration, the effect on the amplitude of the waveform is generally negligible. The increase in pulse power amplitude with bandwidth is shown in Fig. 363. Also shown is the corresponding increase in noise power. The ratio of the two curves indicates the bandwidth for maximum signal-to-noise ratio. This occurs at about $\Delta f = 0.7/d$ for video transmission ($1.5/d$ for double sideband transmission).

59. Bandwidth Criterion for High Resolution.—The foregoing argument has proceeded on the assumption that the amplitude of the transmitted pulse must override the noise level as much as possible, in the interest of discerning the signal. This is the criterion in the vast majority of radar applications. But in certain radar problems, particularly those which involve the measurement of target range to very high accuracy or the discernment of small targets against reflecting background, the *shape* of the output pulse may assume importance.

This is evidently the case when it is desired to measure the reflection interval to a precision equivalent to a fraction of the pulse width. Such measurements must be made against the leading or trailing edge of the pulse, and accordingly the edges

must be as steep as possible. We have seen from Fig. 86 that when the bandwidth extends out only to the first zero of the rectangular pulse spectrum, the reproduced pulse has sides with a definite slope. The pulse shape after transmission is not rectangular but more nearly triangular. If the bandwidth is increased to include the second and successively higher zeroes of the pulse spectrum, the steepness of the edges of the output pulse is increased in proportion to the bandwidth. In radars that must display high resolution, values of bandwidth as high as five times the inverse pulse width ($5/d$) are sometimes used.

When such wide bands are used, the signal-to-noise power ratio suffers as an inevitable consequence. In the case of a bandwidth equal to $4/d$ the signal-to-noise power ratio is reduced about four times or 6 db. The 6 db loss corresponds to a loss of about 30 per cent in the maximum range of the radar, all other factors being unchanged.

Fortunately the loss is not generally as serious as this because the pulse may be distinguished from the noise by its shape as well as by its amplitude. When a wide band is used, the output pulse is nearly rectangular, but the noise output being generated by pulses of infinitesimal duration is composed of a multiplicity of triangular shapes of varying heights. The distinction between the signal and noise is further aided by the steadiness of the signal and the random unsteadiness of the noise. The net effect is that a pulse may be perceived in a wideband radar on a type A indicator, even when its amplitude is somewhat lower than the peak amplitude of the noise. Considerable use of this fact has been made in specialized radars.

60. The Noise Concept in Radar.—The importance of noise in radar design has forced an examination of noise concepts, which has resulted in a reevaluation and redefinition of noise parameters. As late as 1941, the thermal noise contributed by a circuit was taken as

$$P_N = 4kT\,\Delta f \tag{127}$$

where k is Boltzmann's constant (1.37×10^{-23} watt-sec per deg Kelvin), T is the absolute temperature of the circuit in which the noise appears, and Δf is the bandwidth defined, as in Eq. (126), as

$$\Delta f = \int G(f)\,df \tag{128}$$

But in that year, careful measurements proved that the noise power contributed by a circuit was definitely less than this, and attention was focused on the factor "4" in Eq. (127). In 1941 Friis[1] suggested that the difficulty arose from the fact that it was impossible to abstract all the power from a power source, and he suggested that the maximum amount of power that could be so abstracted be called the "*available power*" of the source.

The concept of available power may be developed from Fig. 91. The noise source has a series voltage source e_N and an internal resistance R_N. It is connected to a noise measuring circuit of resistance R_M, which abstracts noise power from the source. The maximum power flows from source to R_M, according to the familiar theorem of elementary electricity, when $R_N = R_M$. Then the current flowing in the series circuit is

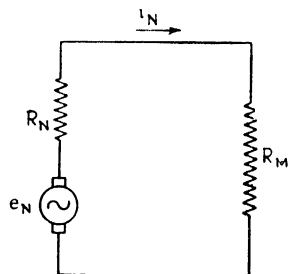



FIG. 91.—Basic circuit for measuring noise.

$$i_N = \frac{e_N}{R_N + R_M} \tag{129}$$

$$= \frac{e_N}{2R_N} \tag{130}$$

The power dissipated in the measuring circuit is then

$$P_M = i_N^2 R_M = \frac{e_N^2}{4R_N^2} R_M \tag{131}$$

$$= \frac{e_N^2}{4R_N} = \frac{1}{4} P_N \tag{132}$$

It thus appears that the maximum available power from the noise source is one fourth the power P_N developed in the source itself.

If we now reexamine Eq. (127), it is true that the *developed* power P_N in a noise source is $4kT\,\Delta f$, but the *available* power P_{NA} of that source is one fourth as much or

$$P_{NA} = kT\,\Delta f \tag{133}$$

This concept was speedily accepted as the basis of radar noise performance.

[1] FRIIS, H. T., Noise Figures of Radio Receivers, *Proc. I.R.E.*, **32**, 419 (July, 1944). The delay in publication resulted from security considerations.

The available power concept is applicable to any source of power, whether of signal or noise. In considering the signal-to-noise power ratio in a radar circuit, the available signal power is compared with the available noise power from all noise sources present. The noise sources in turn are classified as (1) those external to the receiver and (2) those within the receiver itself.

The available noise external to the receiver, originating in the atmosphere and appearing across the radiation resistance of the antenna, is evidently not subject to amelioration by improvements in receiver design. The available external noise power is accordingly a suitable standard for comparison of receiver performance. The available noise power defined in Eq. (133) is taken as the standard of reference. The additional noise introduced within the receiver is measured and compared with $kT\ \Delta f$ to establish the merit of the receiver design with respect to noise.

When Eq. (133) is taken as a standard of reference it is necessary to establish a standard value of T and to measure Δf in accordance with Eq. (128). The value of k is 1.37×10^{-23}, and T is defined for comparative measurements as 290°K. This indicates that the external noise available power is 4×10^{-21} watt per cycle of effective bandwidth.

It should be recognized that the adopted value of T is arbitrary, since the temperature of the radiation resistance of an antenna is not the temperature of the space immediately surrounding it. Rather the value of T represents the average effect of the thermal motion of all charges whose fields act on the antenna. Recent measurements indicate that the proper value of T for radiation resistance is under 100°K. This indicates that the irreducible external noise is about 10^{-21}, or roughly one fourth the value defined for comparative measurements.

The noise generated within the receiver is measured by a standardized procedure, with a thermocouple meter in the output of the receiver. The meter reading is first noted with the input shunted by a resistance equal to the radiation resistance of the intended antenna, but with no signal applied. The meter then reads only the available noise power present in the output. Signal is then applied in series with the input resistance until the thermocouple meter reading is doubled. The available signal power input then present is stated in decibels relative to

$kT\ \Delta f$ as zero db, and it is defined as the *noise figure* of the receiver in question. The noise figure is, in other words, the available signal power input required to double the power (signal plus noise) present in the output, in db relative to $kT\ \Delta f$ as a reference.

The improvement of the noise figure of a receiver depends upon many factors, particularly upon the design of the tubes employed in the preamplifier and mixer stage and in the early i-f stages. These aspects of design are treated in later chapters.

One aspect of circuit design not generally recognized is the improvement of noise figure due to mismatching the impedance of the receiver input to the radiation resistance of the antenna. First consider an ideal receiver having no internal sources of noise except its input impedance, and suppose that the input impedance matches the radiation resistance of the antenna. Then the available noise power is $kT\ \Delta f$ from the antenna and an equal amount from the input impedance. The total noise available power present is then $2kT\ \Delta f$, which is 3 db higher than the reference standard $kT\ \Delta f$. The ideal receiver thus has a noise figure of 3 db, when matched to the antenna. Many radar receivers in the lower frequency ranges (200 megacycles and lower) closely approach this figure.

Suppose, however, that the input resistance is deliberately mismatched to the radiation resistance of the antenna, by connecting the two through loss-less step-up transformer. Then the only noise source present (the ideal receiver again being assumed), is the radiation resistance and the noise figure is unity, or zero db, a 3 db improvement over the matched case. Not all this improvement is actually available, but a measurable improvement is obtainable by mismatching.

Bibliography

GUILLEMIN, E. A.: "Communication Networks," vol. II, John Wiley & Sons, Inc., New York, 1935.

REICH, H. J.: "Theory and Applications of Electron Tubes," McGraw-Hill Book Company, Inc, New York, 1939.

SKILLING, H. H.: "Transient Electric Currents," McGraw-Hill Book Company, Inc., New York, 1937.

FINK, D. G.: "Principles of Television Engineering," Chaps. V and VI, McGraw-Hill Book Company, Inc., New York, 1940.

TERMAN, F. E.: "Radio Engineering," McGraw-Hill Book Company, Inc., New York, 1938.

CHAPTER III

RADIO FREQUENCY FUNDAMENTALS—TRANSMISSION LINES, WAVEGUIDES, AND RESONANT CAVITIES

61. Radio-frequency Techniques in Radar.—In this chapter and the one following we consider the fundamental aspects of r-f (radio-frequency) technology as applied to radar. These aspects include the generation, transmission, radiation, propagation, reflection, and reception of r-f energy.

Radio-frequency energy is generated in resonant circuits or cavities which are part of, or attached to, electron tube structures. The transmission of r-f energy is carried out in transmission lines (open or coaxial) or waveguides. Radiation is accomplished in radiating elements, of which the oscillating dipole is the prototype. The formation of beams involves the use of many dipoles in an array, together with reflector elements, or a continuous reflecting surface in conjunction with one or more dipoles.

Propagation is governed not only by the geometrical factors considered in Chap. I, but by many external influences such as reflection and diffraction of energy at the surface of the earth, refraction in the lower layers of the atmosphere, absorption and scattering in the presence of rainfall, water vapor, etc. Reflection at the target is governed by the type and nature of the discontinuity in the electrical properties between the target and the medium surrounding it. Finally, reception of the r-f signal is governed by the same factors as radiation.

In what follows we consider the general relationships that underlie the design and operation of r-f systems. Particular circuits and r-f structures, including vacuum tubes, are treated in Part II, Chap. VII.

62. Radio Frequencies Employed in Radar.—The radio frequencies employed in radar comprise the entire spectrum from 15 megacycles upwards. The intermediate frequencies in

receivers range from 15 to 60 megacycles. Early radars (the British CH stations) radiated energy at a carrier frequency of approximately 25 megacycles. The loran navigation system, which employs the pulse technique, operates in the vicinity of 2 megacycles, and it is currently being developed to operate at a carrier frequency in the region from 150 to 200 kc.

For the most part, however, attention has been confined in radar work to frequencies above 200 megacycles. Radar development has followed five separate bands of frequencies in this region, which have been given the letter designations listed at the bottom of Table IV.

TABLE IV.—RADIO-FREQUENCY DESIGNATIONS

Designation	Frequency limits mc	Wavelength m
Vlf (very low frequency)	0.010–0.030	30,000–10,000
L-f (low frequency)	0.030–0.300	10,000–1,000
M-f (medium frequency)	0.300–3.0	1,000–100
H-f (high frequency)	3.000–30.00	100–10
Vhf (very high frequency)	30.00–300.0	10–1
Uhf (Ultra high frequency)	300.0–3000.	1–0.1
Shf (Super high frequency)	3,000–30,000	0.1–0.01
	mc	cm
P band	225–390	133–77
L band	390–1,550	77–19.35
S band	1,550–5,200	19.35–5.77
X band	5,200–11,000	5.77–2.73
K band	11,000–33,000	2.73–0.91

In the L, S, X, and K bands, subscripts are used (for example, S_a = 3,100 to 3,400 megacycles) to designate particular regions within each band. The S, X, and K bands are commonly known as the "microwave" or "centimeter-wave" regions.

The most important generalization to be drawn from these figures is the fact that radar waves, from 0.9 cm to 1.33 m in length, are so short that it is possible to build radiating structures whose dimensions extend over many wavelengths. This permits concentration of the energy in a narrow beam with consequent high directivity and high power density. The second generaliza-

tion is that the transmission lines and circuit elements employed also may assume dimensions several wavelengths long. This fact is not always a convenience and in fact requires precautions to avoid unwanted loss or transfer of r-f energy.

Generally speaking the lower frequencies, 200 to 500 megacycles, are used in long-range, ground-based, or shipborne radars to warn of the approach of aircraft and ships at great ranges. The h-f (high-frequency) radars, from 1,500 megacycles to 30,000 megacycles, are used in airborne applications and in gun-directing radars where extreme precision of bearing is required. Microwave equipment is also used for long-range warning. Chronologically speaking, radar development has progressed steadily toward the higher frequencies. The lower frequency equipments, below 500 megacycles, are generally considered obsolete. Principal reliance has been placed on S band (10-cm) and X band (3-cm) equipment for all air-, ship-, and ground-borne uses.

63. Transmission Lines.—A convenient starting point in the study of r-f structures is the transmission line, not only because it is familiar, but because it leads naturally to consideration of waveguides and resonant cavities. The primary function of a transmission line is to convey r-f energy at whatever energy level is required from one point to another with a minimum of loss and without serious discrimination among the several radio frequencies present. Emphasis is thus placed on the losses incident to transmission and on the maximum power-handling capability of the line.

The losses in transmission lines are of three basic types: (1) loss in heating the transmission line conductors, the so-called "i^2r" loss. This loss is dependent on the resistivity of the conducting material and the shape of the conducting paths. The skin effect causes h-f currents to flow in a thin sheath of each conductor, so the perimeter of the cross section of the line is the significant geometrical factor affecting i^2r loss. (2) Loss due to radiation. This is present in appreciable amounts only in open-wire transmission lines when standing waves are present. (3) Loss in the insulation between conductors. This loss depends upon the quality and amount of the insulation. When standing waves are present, the position of the insulation, relative to the voltage maxima, is also a criterion.

All these losses are considerably increased when standing waves are present, because of the higher currents and voltages present as well as the multiple traverses of the power along the line. Hence a very important aspect of transmission line engineering is the reduction of standing waves by proper termination and adjustment of the line.

It is not difficult to design a *uniform* transmission line that will pass a wide band of frequencies without discrimination, provided that the dimensions of the line can be chosen properly with respect to the middle frequency of the band. But a truly uniform transmission line is almost impossible to achieve in practice, particularly when resonant elements are used as supports. When such resonances are present, an additional problem arises, that is, broadening the range over which frequencies can be transmitted without undue discrimination.

The two general types of line are the *open-wire* or *parallel-wire* line and the *coaxial* or *concentric* line. Early in radar development the two types were in competition. The open-wire line was used at radio frequencies where high-power capability and low loss are paramount and the coaxial line for pulse (video-frequency) transmission where shielding is the first consideration. But it has since proved feasible to employ coaxial lines or waveguides in all applications where open lines were once thought essential. The coaxial line is preferred because of the ease with which it can be put in place and moved and its freedom from extraneous electrical influences. At the highest frequencies, moreover, the open line is ruled out by the very small separation (one-tenth wavelength or less) required between wires to minimize radiation.

64. Mechanism of Transmission.—The propagation of r-f energy along a transmission line may be considered from either of two points of view. The first approach, the more widely understood among engineers, is to consider the line as a circuit composed of many small elements of resistance, inductance, and capacitance, and to allow the size of these elements to approach zero while their number increases to infinity. These are the limits at which the R, L, and C elements are smoothly distributed along the line. The second approach, which derives from the treatment of the waveguide, is to consider the transmission line as a pair of conducting boundaries which guide the r-f energy

electromagnetically in accordance with the Maxwell equations. In this latter approach it is found that there are many configurations or *modes* by which the energy may be transmitted along the line, only one of which corresponds to the result obtained from the analysis by distributed parameters. This mode (the principal mode) is the only mode that can exist physically when the cross-sectional dimensions of the transmission line are small compared to the wavelength transmitted. By proper choice of the dimensions of the line, therefore, it is possible to restrict the mode of transmission to the principal mode, as is almost always done in practice. At the higher frequencies, the transmission line becomes impractically small under this restriction, and therefore waveguides are used.

FIG. 92.—Representation (*a*) of voltage and current on a uniform transmission line, and corresponding lumped-parameter circuit (*b*).

We consider here the distributed parameter case. Figure 92 shows a two-conductor line (which may represent either an open-wire line or a coaxial line), along which sinusoidal r-f energy is traveling. Distance x is measured along the line from the left, and the voltage across the line V and the current through it I are taken as the rms amplitudes at the distance x from the left-hand end. The line itself is represented electrically by a series impedance

$$Z = R + j\omega L \qquad \text{ohms per m} \tag{134}$$

and a shunt admittance

$$Y = G + j\omega C \qquad \text{mhos per m} \tag{135}$$

Here R is the loop resistance (two conductors in series) of the line per meter length of line, L the loop inductance per meter, G the conductance (insulation losses) per meter, and C the shunt capacitance per meter.

Consider the change in voltage dV and the change in current dI that occur in a very small (differential) length dx of the line.

The voltage drop is equal to the impedance per meter length times the current times the distance element dx. The current change is equal to the admittance per meter length times the voltage across the line at that point, times dx. Hence

$$\frac{dV}{dx} = ZI \tag{136}$$

and

$$\frac{dI}{dx} = YV \tag{137}$$

It is now required to find the voltage V and the current I as functions of x, Z, and Y. The equations are solved by taking the second derivative of each and substituting, as follows:

$$\frac{d^2V}{dx^2} = \frac{Z\,dI}{dx} = ZYV \tag{138}$$

$$\frac{d^2I}{dx^2} = \frac{Y\,dV}{dx} = ZYI \tag{139}$$

It is evident that the solution for the voltage and that for the current have the same form. Let us consider the voltage equation. We assume that the voltage will vary in accordance with the form

$$V = V_1\epsilon^{+\sqrt{ZY}x} + V_2\epsilon^{-\sqrt{ZY}x} \tag{140}$$

where $\epsilon = 2.718$ is the base of the Napierian logarithms, and $V_1 + V_2$ is the voltage across the line at $x = 0$. Substitution of Eq. (140) in Eq. (138) results in an identity, so Eq. (140) is, in fact, a solution. Equation (140) reveals that the rms amplitudes of the sinusoidal voltages on the line vary logarithmically with the distance and with the quantity $\sqrt{ZY}$, the geometric mean of the series impedance and shunt admittance of the line. The larger Z or Y is, the more rapid is the attenuation of the signal with distance.

The plus and minus signs in the exponents of the two terms of Eq. (140) indicate that the signal may be attenuated either to the right or to the left, with increasing or decreasing x. The physical interpretation is that the line may transmit two waves at the same time, one attenuated to the left, the other to the right. One of these waves is referred to as the "direct wave." The other, a reflection of the first, is the "reflected wave."

Similarly the current in the line is given by

$$I = I_1\epsilon^{+\sqrt{ZY}x} + I_2\epsilon^{-\sqrt{ZY}x} \qquad (141)$$

The quantity $\sqrt{ZY}$ is complex.

$$\sqrt{ZY} = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta \qquad (142)$$

where α, the real part, is known as the attenuation constant, and β, the imaginary part, is known as the phase constant. The expressions for α and β are complicated when the effects of R, L, C, and G are equally prominent in the line. In practice, at high frequencies, the terms ωL and ωC far exceed R and G, respectively. Then the approximate expressions for α and β are

$$\alpha = \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \qquad (143)$$

and

$$\beta = \omega\sqrt{LC} \qquad (144)$$

The attenuation constant governs the rate at which the signal amplitude decreases with distance along the line, whereas the phase constant determines the wavelength along the line. The physical situation is as follows: if the wave has a given amplitude at a given point, then at a distance $1/\alpha$ m further along the line the amplitude has been divided by 2.718. If the wave has a positive maximum at a given point, the next positive maximum occurs $2\pi/\beta$ m further along the line. These relationships are shown in Fig. 93.

A most important fact, expressed in Eq. (143), is that the attenuation of the signal is independent of frequency, at least to the extent that R, L, C, and G do not vary with frequency. In other words, the line is inherently a "wide band" device. The series resistance R does in fact increase with the frequency due to skin effect, and thus there is always some increase in the attenuation as the frequency is extended indefinitely upward. But the attenuation is essentially constant over the frequency ranges encountered in r-f bandpass spectra. Similarly, Eq. (144) shows that the phase shift introduced by the line is proportional to the angular frequency ω. This is equivalent to constant time delay at all frequencies. It is thus evident that the uniform transmission line is ideally suited to the transmission of pulse

waveforms without distortion of their shape, either at pulse frequencies or at (modulated) radio frequencies.

The phase constant β indicates the velocity at which the wave moves along the line. The wavelength of the transmitted signal,

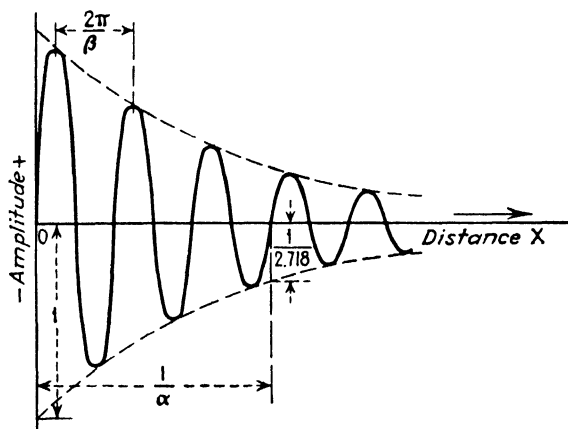



FIG. 93.—Wave on a transmission line, showing significance of attenuation constant α and phase constant β.

as it exists in the line, is

$$\lambda = \frac{2\pi}{\beta} \tag{145}$$

$$= \frac{2\pi}{\omega\sqrt{LC}} = \frac{1}{f\sqrt{LC}} \tag{146}$$

Hence

$$c_L = \lambda f = \frac{1}{\sqrt{LC}} \tag{147}$$

The quantity c_L, the velocity of propagation along the line, is thus shown to be numerically equal to $1/\sqrt{LC}$. When the line is immersed in air, the inductance and capacitance are so related that $1/\sqrt{LC}$ is very nearly equal to the velocity of light. If the line is "loaded" by the insertion of lumped inductance or capacitance or if it contains substances of high permeability or dielectric constant, L and C are increased and the velocity of propagation is decreased in accordance with Eq. (147).

It should be noted that the velocity c_L is the speed with which a given phase of the waveform propagates itself. Hence it is called the "phase velocity." In the case of the line here considered,

operated in its principal mode, the phase velocity is also equal to the velocity at which the energy of the wave is transmitted. In waveguides operated at high modes a distinction between these two velocities must be drawn, as shown in Sec. 75.

65. Characteristic and Terminating Impedances.—A matter of evident importance is the relative prominence of the direct and reflected waves. To secure maximum power transmission through the line, it is necessary to reduce the reflected wave as much as possible. Moreover, when short pulses are transmitted over a line, the reflections introduce additional pulse signals that may interfere with the operation of the radar.

The reflected wave is a minimum (theoretically zero) when the impedance connected across the output termination of the line is equal to a particular value, known as the "characteristic impedance" of the line. The characteristic impedance is the ratio of the d-w (direct-wave) rms voltage amplitude across the line at any point to the d-w rms current amplitude at the same point.

To determine the characteristic impedance, we consider the direct wave of current given in Eq. (141)

$$I = I_1 \epsilon^{+\sqrt{ZY}x} \tag{148}$$

differentiate with respect to x and substitute in Eq. (138)

$$YV = \frac{dI}{dx} = \sqrt{ZY}\, I_1 \epsilon^{+\sqrt{ZY}x} \tag{149}$$

For V in the above we substitute the corresponding direct wave of voltage from Eq. (140) and obtain

$$YV = YV_1 \epsilon^{+\sqrt{YZ}x} = \sqrt{ZY}\, I_1 \epsilon^{+\sqrt{ZY}x} \tag{150}$$

The characteristic impedance is then

$$Z_0 = \frac{V_1}{I_1} = \sqrt{\frac{Z}{Y}} \qquad \text{ohms} \tag{151}$$

In terms of the line parameters, Eqs. (134) and (135), the characteristic impedance is

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \qquad \text{ohms} \tag{151a}$$

When R and G are appreciable compared with the reactances ωL and $1/\omega C$, the characteristic impedance is a function of frequency.

But in a well-designed line, R and G are negligible at operating frequencies. Then

$$Z_0 = \sqrt{\frac{L}{C}} \quad \text{ohms} \tag{151b}$$

which is independent of frequency and purely resistive.

The same result is obtained when the rms amplitudes of the reflected waves V_2 and I_2 are considered Thus in both the

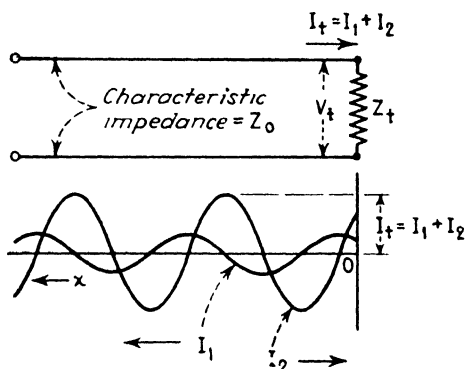



FIG 94 —Reflected wave at termination of a transmission line

direct and reflected waves, considered separately, the voltage and current amplitudes are attenuated together in such a way that their ratio at any point is equal to the characteristic impedance

To determine the extent of the reflected wave, let us extend the transmission line to its output as shown in Fig 94, and connect a terminating impedance

$$Z_t = R_t + jX_t \quad \text{ohms} \tag{152}$$

We now inquire what is the effect of the terminating impedance on the voltage and current waves along the line We measure distance x from the output termination backwards toward the input, and consider the rms voltage and current amplitudes V_t and I_t at the output termination On the one hand, the termination voltage must equal the terminating impedance times the sum of the direct and reflected current amplitudes

$$V_t = Z_t(I_1 + I_2) \tag{153}$$

On the other hand this voltage is, by Eq. (137)

$$V_t = \frac{(dI/dx)_0}{Y} \tag{154}$$

where $(dI/dx)_0$ is the rate of change of current at $x = 0$. Differentiating Eq. (141) and inserting $x = 0$ we obtain

$$V_t = \sqrt{\frac{Z}{Y}}(I_1 - I_2) = Z_0(I_1 - I_2) \tag{155}$$

Equations (153) and 155) are then solved simultaneously to

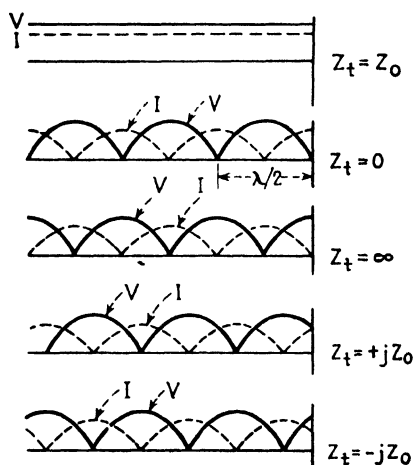



FIG. 95.—Standing waves of current and voltage corresponding to various terminating impedances.

give the relationship between the two current amplitudes.

$$\frac{I_2}{I_1} = \frac{Z_0 - Z_t}{Z_0 + Z_t} \tag{156}$$

Thus, when the characteristic impedance and the terminating impedance are known, it is possible to compute the ratio of the reflected wave I_2 to the direct wave I_1 (the subscript "2" is still associated with the reflected wave although we have taken x increasing to the left). In particular when the line is terminated with an impedance equal to its characteristic impedance, $Z_0 = Z_t$, by Eq. (156), $I_2/I_1 = 0$. Hence no matter what the amplitude of the direct wave I_1 may be, the reflected wave has zero amplitude, that is, there is no reflection. The sum of

the direct and reflected waves for various values of Z_t is shown in Fig. 95.

Finally, we consider the impedance looking into a terminated line of any length, say l m. Then at $x = l$ the voltage is, by analogy to Eq. (155)

$$V_l = Z_0(I_1\epsilon^{-\sqrt{ZY}l} - I_2\epsilon^{+\sqrt{ZY}l}) \tag{157}$$

and the current is

$$I_l = I_1\epsilon^{-\sqrt{ZY}l} + I_2\epsilon^{+\sqrt{ZY}l} \tag{158}$$

The impedance is the ratio V_l/I_l which, by combining Eqs. (156), (157), and (158), can be written

$$Z_l = Z_0\frac{Z_t + Z_0 \tanh \sqrt{ZY}\,l}{Z_0 + Z_t \tanh \sqrt{ZY}\,l} \tag{159}$$

where than $\sqrt{ZY}\,l$, the hyperbolic tangent, is

$$\tanh \sqrt{ZY}\,l = \frac{\epsilon^{\sqrt{ZY}l} - \epsilon^{\sqrt{ZY}l}}{\epsilon^{\sqrt{ZY}l} + \epsilon^{\sqrt{ZY}l}} \tag{160}$$

Thus far we have treated the terminated line in terms of impedances. The symmetry of the equations with respect to Z and Y shows that when an admittance is substituted for an impedance the derivation proceeds without change. Hence, the input admittance of a line l m long, terminated in Y_t admittance, and having Y_0 characteristic admittance has the form of Eq. (159).

$$Y_l = Y_0\frac{Y_t + Y_0 \tanh \sqrt{ZY}\,l}{Y_0 + Y_t \tanh \sqrt{ZY}\,l} \tag{161}$$

Equations (159) and (161) are of great value in deducing the impedance and admittance properties of terminated lines.

66. Loss-free Lines.—Having established the general admittance and impedance relationships of terminated lines, we turn now to a simplification which is admissible in practice and which leads to correspondingly useful results. In the uhf (ultrahigh-frequency) and shf (superhigh-frequency) regions the transmission lines employed are generally several wavelengths long. The resistance and conductance, while present in sufficient amount to account for a definite attenuation along the line, are nevertheless so small compared with the reactive components that

the line may be considered for practical purposes loss free. Then the attenuation factor α may be neglected and the general input impedance relationship Eq. (159) reduces to

$$Z_l = Z_0 \frac{Z_t + jZ_0 \tan \beta l}{Z_0 + jZ_t \tan \beta l} \tag{162}$$

The admittance expression is precisely the same, Y being substituted wherever Z appears. The substitutions by which Eq. (162) is arrived at will be evident from Eqs. (142) and (159).

A number of important properties are revealed by Eq. (162). First we note that the loss-free line, in common with the loss-

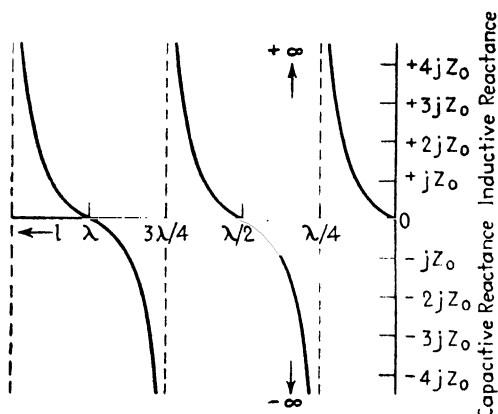



FIG. 96.—Variation of input impedance with length of a transmission line short-circuited at the far end.

full (attenuating) line, has an input impedance equal to its characteristic impedance whenever the terminating impedance is equal to the characteristic impedance. This becomes evident by substituting $Z_t = Z_0$ in Eq. (162), whereupon $Z_l = Z_0$.

As a second case consider that the line is short-circuited at its output, that is, $Z_t = 0$. Then, by Eq. (162)

$$Z_l = jZ_0 \tan \beta l \tag{163}$$

The j in this equation reveals that the input impedance of a short-circuited loss-free line is always reactive, never resistive. Moreover the reactance presented varies cyclically, in accordance with the tangent function, as the length of the line l is increased. The variation of impedance with length is shown graphically in Fig. 96.

When the length of the short-circuited line is some multiple of a quarter wavelength, particularly important cases occur.

TABLE V.—INPUT IMPEDANCE OF LOSS-FREE LINES

Length, wavelengths	Terminating impedance—Z_t	
	Short-circuit	Open-circuit
0	0	∞
⅛	$+jZ_0$	$-jZ_0$
¼	∞	0
⅜	$-jZ_0$	$+jZ_0$
½	0	∞
⅝	$+jZ_0$	$-jZ_0$
¾	∞	0
⅞	$-jZ_0$	$+jZ_0$
1	0	∞

To investigate these cases, we remember from Eq. (145) that $\beta = 2\pi/\lambda$ and $\beta l = 2\pi l/\lambda$. Hence βl, the argument of the tangent in Eq. (163), is the length of the line, expressed in electrical radians. If we consider a short-circuited line one-quarter wavelength long, then $\beta l = \pi/2$ and $\tan \beta l = \infty$. We thus arrive at the conclusion that the short-circuited quarter-wave loss-free line has an infinite input impedance. This property is much used to supply "insulation" at radio frequency without the use of power-consuming and fragile dielectric materials. Similarly a short-circuited halfwave line implies $\tan \beta l = \tan \pi = 0$. The input impedance is then equal to the short circuit, that is, zero. These and other properties of the short-circuited line are illustrated in the second column of Table V.

A closely related case is that of the open-circuited loss-free line $Z_t = \infty$. Then Eq. (162) becomes

$$Z_l = -jZ_0 \cot \beta l \tag{163a}$$

Since the cotangent is the inverse of the tangent, it follows that the input impedance of the open-circuit line is opposite in sign to that of the short-circuited line. Also the infinite and zero input impedances change places, as shown in Table V.

The open-circuited line (third column of Table V) is capable of all the functions of the short-circuited line but it is not much used because its shielding is incomplete and the maintenance of high impedance across an open-ended line is a difficult practical matter.

Returning again to Eq. (162) we note that the half-wavelength line has the following property ($\tan \beta l = 0$)

$$Z_l = Z_0 \frac{Z_t + 0}{Z_0 + 0} = Z_t \qquad (164)$$

Thus the input impedance of a loss-free half-wave line is equal to the terminating impedance. The half-wave line is, in other words, an ideal one-to-one impedance transformer, for which practical uses are occasionally found. The most important practical implication is that a line may be extended any number of half wavelengths without affecting the input impedance, so long as the losses may be neglected.

The one-eighth wavelength line is another case of practical interest because it provides reactance, when short-circuited or open-circuited, the magnitude of the reactance being equal to the characteristic impedance. The reactance is positive (inductive) in the short-circuited case and negative (capacitive) in the open-circuited case. The inductive and capacitive effects exchange in the three-eighths wavelength loss-less line. These properties are listed in Table V.

A final and most important property revealed by Eq. (162) is that of the quarter-wavelength line terminated in the arbitrary impedance Z_t. Taking $\beta l = \pi/2$ ($\tan \beta l = \infty$), Z_l becomes

$$Z_l = Z_0 \frac{jZ_0 \infty}{jZ_t \infty} = \frac{Z_0^2}{Z_t} \qquad (165)$$

This may be rewritten as

$$Z_0 = \sqrt{Z_l Z_t} \qquad (166)$$

This equation states that reflections are absent when the characteristic impedance of a loss-free quarter-wave line is the geometric mean (square root of the product) of the terminating impedances. Thus it is possible to couple two unlike impedances, so long as both are resistive, by inserting a quarter-wave section whose characteristic impedance meets the condition of Eq. (166).

The quarter-wave impedance transformer is basic to impedance matching procedure in all radar r-f circuits and structures.

The foregoing properties of loss-free lines apply with sufficient accuracy for engineering purposes when the lines display appreciable losses, so long as the frequency of use is high enough to make the attenuation constant α small compared with

$$\beta = \omega \sqrt{LC}.$$

This condition is met in nearly all radar r-f applications.

67. Impedance and Admittance Circle Diagrams.—The generalities drawn in the preceding section concern loss-free lines of particular lengths. If we wish to examine the input impedance of *any* length of line with *any* termination, we must return to the general impedance expressions for the loss-full line Eq. (159) and the loss-free line Eq. (162) and the corresponding admittance expressions. These expressions are cumbersome to handle numerically, but they can be expressed conveniently in graphical form, in rectangular coordinates. The resistive element of the impedance is plotted horizontally as abscissas and the reactive element vertically as ordinates. For the time being we shall confine our attention to the impedance diagram.[1]

The impedance diagram comprises a group of loci of points representing the resistive and reactive parts of the line input impedance. The loci are found to be circular in form and hence the term "circle diagram." By the use of the circle diagram, it is possible to solve all manner of transmission line problems, among which are the following:

1. Determining the input impedance for any length of line, with any value of characteristic impedance, for any value of terminating impedance. Loss-full and loss-free cases may be treated.
2. Determining the relative amplitudes of the direct and reflected waves at any discontinuity in the line.
3. Determining the effect on input impedance of series impedances of whatever form inserted in the line. Shunt admittances may be similarly treated with the admittance circle diagram.
4. The design of impedance matching stubs and transformers.

[1] The admittance diagram, which has the same shape (conductance horizontally and susceptance vertically) but a slightly different rule of application, is discussed in Sec. 69.

5. The design of step-by-step impedance matching devices having wide-band properties.

It is evident from this list that the impedance and admittance circle diagrams are very important tools in r-f transmission line engineering. Since their advent is recent in communications practice, it is urged that the reader acquire a working familiarity with such charts if he does not already possess it.

There are two widely used forms of circle diagrams. The first is simply a resistance-reactance diagram with circles plotted on

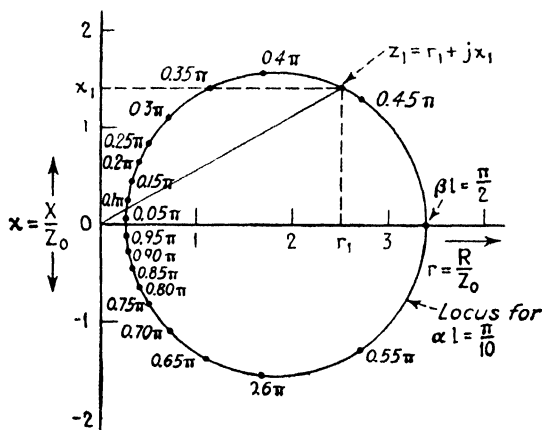



FIG. 97.—Representation of transmission line normalized impedance in terms of constant-αl and constant-βl loci (rectangular form of circle diagram).

it by direct reference to Eqs. (158) or (162). The other, a transformation of the first devised by Smith,[1] is somewhat easier to use and gives standing-wave information directly. Both forms are considered below.

Consider first the resistance-reactance plot, in rectangular coordinates, shown in Fig. 97. The resistance and reactance are plotted as normalized fractions of the characteristic impedance, $r = R/Z_0$ and $x = X/Z_0$, so that the charts may apply to transmission lines having any value of Z_0. It is assumed that the characteristic impedance is purely resistive which is quite accurately true of all practical transmission lines (except those constructed to have abnormally high losses for special applications). A point r_l, x_l on the diagram defines a vector impedance,

[1] SMITH, P. H., Transmission Line Calculator, *Electronics*, **12**, 29 (January, 1939).

drawn from the origin to the point, and this impedance $(r_l + jx_l)$ is a particular normalized value $z_l = Z_l/Z_0$ of the impedance looking into the input end of the transmission line in ques-

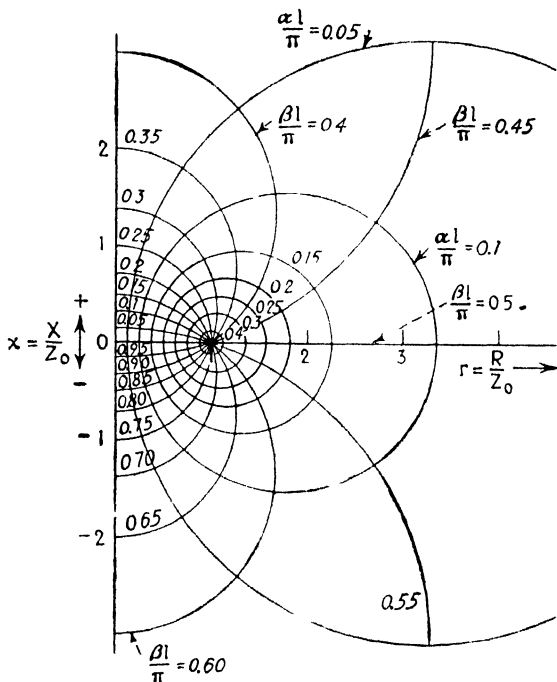



FIG. 98.—Circle diagram showing several constant-αl and constant-βl loci.

tion. Rewriting Eq. (159) in terms of α and β, we have for this impedance[1]

$$z_l = \frac{Z_l}{Z_0} = \frac{\sinh \alpha l \cosh \alpha l + j \sin \beta l \cos \beta l}{\cosh^2 \alpha l \cos^2 \beta l + \sinh^2 \alpha l \sin^2 \beta l} \tag{167}$$

If we choose some particular value of αl in Eq. (167) and allow the quantity βl to take on all values from zero to π, we find that the real and imaginary parts of z_l define points that fall on a circular locus, such as is shown in Fig. 97. The corresponding values of βl around the circle are marked. It will be noted that these values of βl are not evenly distributed around the

[1] The equivalence of this expression to Eq. (159) is not obvious, but may be proved readily by manipulation of the natural and hyperbolic trigonometric quantities. A complete treatment is given by J. C. Slater, "Microwave Transmission," p. 29, McGraw-Hill Book Company, Inc., New York, 1942.

circle, being spread widest in the region of the circle away from the origin. If we choose other values of αl and repeat the process for each, we obtain a family of circles that surround the point $r = 1$, $x = 0$. The circles are not concentric, but have centers that approach the point 1,0 as the circles become smaller.

If we now direct our attention to the values of βl marked on the circles we find that loci may be drawn through equal values of this quantity, as shown in Fig. 98. These loci, also, are circles, and they are orthogonal (cross at right angles) to the first set. The circles of the second set pass through the point 1,0

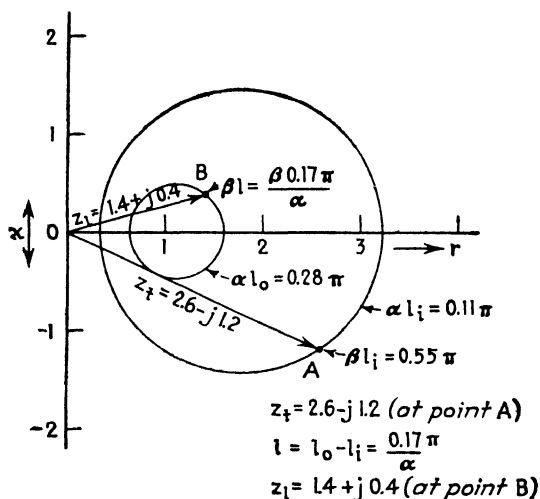



FIG. 99.—Construction for finding the input impedance z_l of a given length of line l_0 terminated in a given impedance z_t.

and center on the x axis. These two families of circles, with the rectangular coordinates as a background, present all possible values of attenuation constant α, phase constant β, line length l, and terminating impedance $z_t = r_t + jx_t$. In treating a given transmission line, having particular values of α, β, l, r_t, and x_t, a point is found on the chart that satisfies these values simultaneously. This point specifies a normalized resistance r_l and reactance x_l, which are the components of the normalized input line impedance under the given conditions.

To illustrate the process, consider Fig. 99. The terminating impedance z_t is shown as the point A. From Fig. 98 we find the corresponding value of αl of the circle passing through this

point and construct the circle. This value of αl is divided by the attenuation constant α of the given line, found by Eq. (143), producing a particular value of l, denoted by l_t. This length represents zero length of transmission line, since it occurs at the terminating impedance. To find the value of input impedance of a line of length l_0 m we find β from Eq. (144) and form the quantity $\beta l = \beta(l_0 - l_t)$. We then locate the value of βl on the chart where it intersects the corresponding value of αl_0. The

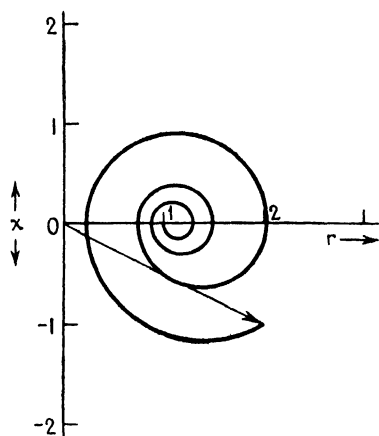



FIG. 100.—Spiral representing input impedance as a function of length of a terminated loss-full line.

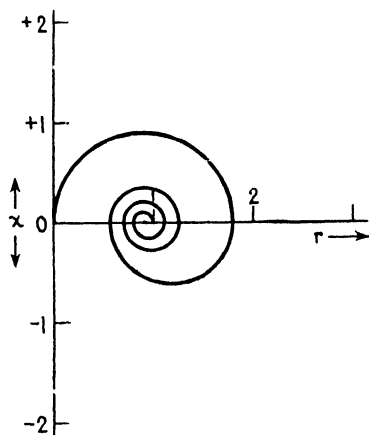



FIG. 101.—Impedance spiral of short-circuited loss-full line.

point of intersection B gives the components r_l and x_l of the input impedance.

In addition to this straightforward method of determining input impedances under any condition of termination, the circle diagram is very useful as an aid in visualizing the change in input impedance as the length of line is indefinitely increased. Consider Fig. 100, a spiral representing the variations in impedance in a loss-full line. The most counterclockwise point of the spiral represents the components of the terminating impedance. As we increase the length of the line, the increase in the quantity βl causes us to proceed clockwise around the spiral. Each full turn in the spiral, as reference to Fig. 98 shows, represents a half wavelength of line. The spiral is formed from the con stant αl circles of Fig. 98, the radius of the circle continually decreasing as the quantity αl increases with the increasing

length of line. The spiral finally winds up around the point $r = 1$, $x = 0$, that is,

$$z_l = \frac{Z_l}{Z_0} = 1 \tag{168}$$

from which

$$Z_l = Z_0 \tag{169}$$

Hence no matter how a loss-full line is terminated, as its length

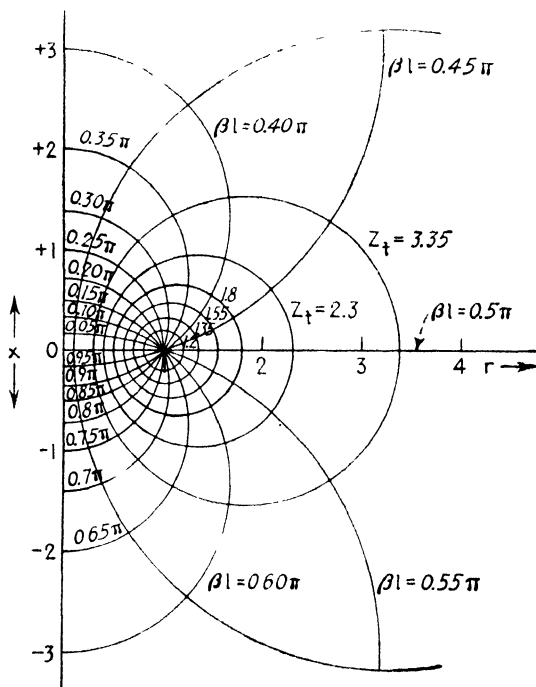



FIG. 102.—Rectangular circle diagram of the loss-free line, in which loci of constant terminating impedance z_t are plotted in place of the constant-αl loci of Fig. 98.

is increased the input impedance approaches the characteristic impedance. A special case of the spiral is shown in Fig. 101 for the short-circuited line, which starts at the point 0,0 and proceeds clockwise.

68. Circle Diagram of the Loss-free Line.—In many cases of practical interest, the attenuation constant α is so small compared with the phase shift constant β that the spirals of Figs. 100 and 101 degenerate into circles.

The circle diagram in this case is derived from the loss-free input impedance equation, Eq. (162). We plot circles of constant normalized terminating impedance $z_t = Z_t/Z_0$ and allow βl to assume all values from 0 to π, as before. The constant Z_t circles surround the point 1,0, and corresponding values of βl can be joined by circles orthogonal to the first, the second set of circles centering on the x axis. The diagram, Fig. 102, has precisely the same form as Fig. 98.

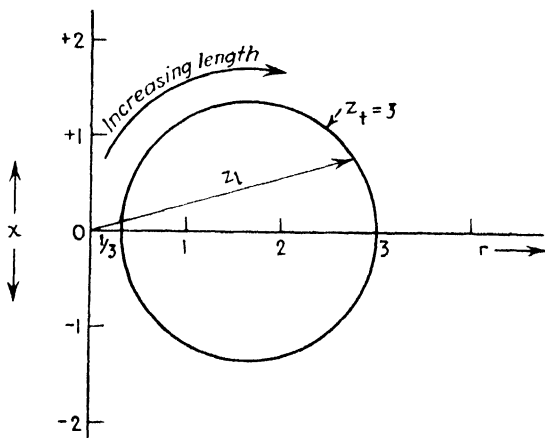



FIG. 103.—Construction of constant-z_t circle for loss-free line and relationship to input impedance z_l.

As an example, in Fig. 103, the line is terminated by a resistance three times the characteristic impedance, that is, $z_t = 3$. Then the impedance looking into the line will follow around the $z_t = 3$ circle as the line length is increased. Zero length is indicated by the point on the r axis farthest from the origin, where the impedance is evidently $z_l = z_t$. One-quarter wave-length is represented on the r axis at the intersection nearest the origin. The value of impedance at this point is

$$z_l = \frac{1}{z_t} \tag{170}$$

This follows from the fact the two impedances separated one-quarter wavelength must, by Eq. (166), have a geometric mean equal to the normalized characteristic impedance (unity). Hence they must have reciprocal values. This gives us a convenient rule for constructing the circles. The outer intersection

must pass through the associated value z_t, and the inner intersection must pass through the reciprocal $1/z_t$. Hence the radius of the circle is

$$r = \frac{z_t - 1/z_t}{2} \tag{171}$$

The orthogonal circles, which connect equal values of βl, may be constructed by using a radius $r = \tan \beta l$, passing through the point 1,0 and centered on the x axis.

This form of diagram may be used in many design problems, of which that illustrated in Fig. 104 is typical. Consider two transmission lines, terminated in such ways and of such lengths that

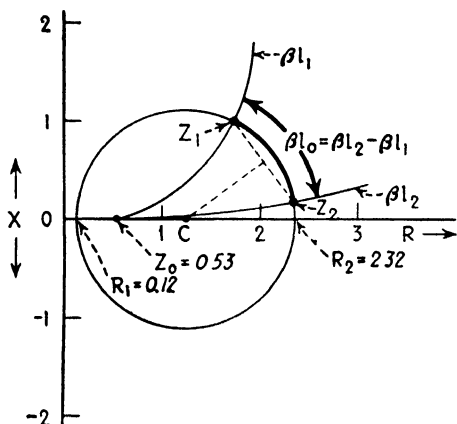



FIG. 104.—Construction for finding characteristic impedance and length of matching section between two dissimilar lines.

they present impedances Z_1 and Z_2, respectively, at their input ends. Both impedances have reactive as well as resistive components. We are required to join these input ends with a section of line that will match the one input impedance to the other, that is, we must specify the characteristic impedance and the length of the matching section. To do so we plot Z_1 and Z_2 in R,X coordinates, as shown (normalized coordinates are not convenient in this case), and pass a circle centered on the R axis through Z_1 and Z_2. We note the values of resistance at the intersections of the circle with the R axis, R_1, and R_2 as shown. The characteristic impedance required in the matching section is then, by Eq. (166),

$$Z_0 = \sqrt{R_1 R_2} \tag{172}$$

and the length required in the matching section is represented by the portion of the circle, measured clockwise from Z_1 to Z_2. To find the corresponding values of βl, we pass circles through Z_0 and Z_1 in one case and through Z_0 and Z_2 in the other, both circles being centered on the X axis. The radii of these circles are $\tan \beta l_1$ and $\tan \beta l_2$, respectively. From these we can find $\beta l_2 - \beta l_1$, which is the required length of the matching section in electrical radians.

Another basic use of the loss-free circle diagram is illustrated in Fig. 105. Consider a line of characteristic impedance Z_0,

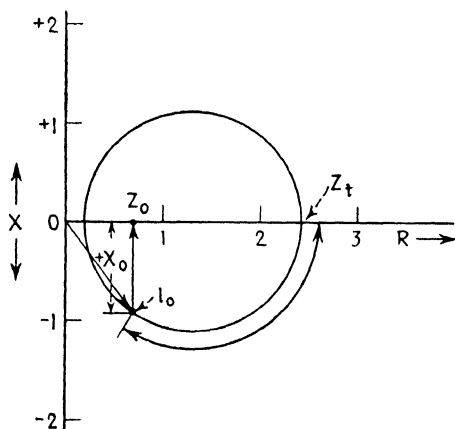



FIG. 105.—Cancellation of reactive component by insertion of series reactance of opposite sign.

terminated in another, resistive, impedance of different value Z_t. Then the circle diagram is that shown, intersecting the R axis at Z_t. Since the line is not terminated in its characteristic impedance, reflections are present. Suppose it is desired to remove these reflections by matching impedances. To do so, we move along the line from the termination to the point represented by l_0, just below the Z_0 value. At this point the resistive component of the input impedance is equal to the characteristic impedance, and the reactive component is $-X_0$ as shown. This reactive component may be removed by inserting at this point in series with the line a reactive length of line. The inserted inductive reactance component (which may be an accurately proportioned short length of line) is represented by a vertical line of length $+X_0$. When a reactance of this value is added

to the line at length l_0, the input impedance is translated to the point Z_0 on the R axis. Thus the input impedance of line with the inductive reactance element present is the characteristic impedance of the line. No reflections can exist on a line that presents its characteristic impedance at its input, and thus the reflections are removed. This method is not much used in practice because of the critical length of the reactive section. Shunt admittance can perform the same function with considerably greater ease of adjustment, as is shown below.

69. The Admittance Circle Diagram and Shunt Reactive Elements.—*In practice it is found difficult, mechanically, to adjust* the length of a line for the purpose of reducing reflections. A simpler method is to employ side stubs of adjustable length. Such stubs present shunt reactance to the line, and hence are most conveniently treated in the admittance terminology, in which shunt elements add directly. This fact establishes the need for an admittance form of the circle diagram.

Fortunately the impedance diagram will suffice for admittances, when proper attention is paid to the sign of the susceptance component. Thus, in accordance with accepted engineering usage, inductive susceptance is a positive quantity.

$$B_L = + \frac{1}{\omega L} \tag{173}$$

Then the admittance becomes

$$Y = G - jB_L \tag{174}$$

Thus if we are to use the admittance diagram without change, we must measure inductive susceptance below the r axis and capacitive susceptance above the axis.[1] This interchanges the positions of inductive and capacitive elements, but in all other respects the diagram and its method of use are unchanged. The length of line, proceeding away from the termination, is measured clockwise.

When a single shunt-reactance stub is added to a transmission line to remove reflections, the diagram (Fig. 106) has exactly

[1] If the inductive elements are measured above the axis, this direction is reversed and length along the line is read counterclockwise. The method used, while optional, should be clearly in mind to avoid errors in admittance matching problems. In the examples given here, we adopt the clockwise motion in all cases and measure inductive susceptance below the r axis.

the same form as for series reactance (Fig. 105). Two cases are shown. In one the stub is less than a quarter wave long (inductive) and in the other more than a quarter wave (capacitive). To remove reflections on a line of characteristic admittance Y_0 the line is broken at a distance measured from the termination corresponding to a point directly above or below Y_0, as shown, and the stub adjusted in length until its susceptance is just equal and opposite to the reactance represented by the vertical distance from that point to the R axis. This addition brings the line input admittance impedance to the characteristic value. There can be no reflections on a line presenting characteristic admittance

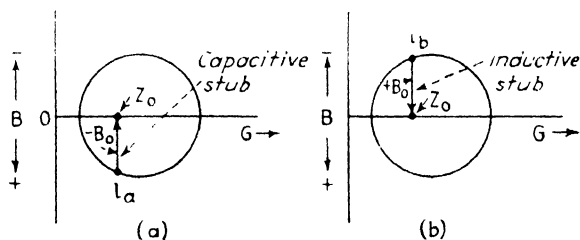



FIG. 106.—Cancellation of susceptance in lines by insertion of shunt susceptance of opposite sign, using the admittance circle diagram.

at its input end. It follows that there is no reflected wave. In practice the stub is adjusted, while the reflections present are detected by measurement of standing waves, until the reflections are a minimum.

The single stub suffers from the fact that it must be placed at some particular point in the line, represented by points l_a and l_b in Fig. 106. To assure that the position is found, it is usually necessary to adjust the length of the line, which as previously stated is inconvenient mechanically. To avoid this difficulty two stubs may be placed in the line at any two fixed points. The preferred separation between them is an odd number of eighth wavelengths, say three-eighths, since this provides a range of impedance adjustment from 0 to twice the characteristic admittance. Still another type of tuner is the triple stub tuner with quarter-wave separation that will match any impedance to any other impedance and is simple to adjust. These and other transmission line tuning elements are treated in Chap. XII.

All the above circle diagrams are computed for one particular

wavelength. If several wavelengths are present (as in modulated transmission) or if the system must be tuned to different frequencies, the points are rotated about the constant-x_t circles by the fact that the length of the line changes, in electrical radians, proportionately to the frequency. Thus a given set of conditions, represented by a circle diagram at a particular frequency, becomes quite different at another frequency. This represents the selectivity of the tuning elements employed. For example, quarter-wave transformers and matching stubs change their length, electrically, with the frequency and minimum reflection (maximum power transmission) is achieved only at the frequency for which the system has been adjusted. To decrease the selectivity of matching elements, the change in characteristic impedance at each junction must be made as gradual as possible, by tapering the line dimensions or by introducing graduated changes in the line parameters.

70. The Smith Circle Diagram.—A particularly convenient form of the circle diagram, much used in practice, is that shown in Fig. 107. This form of circle chart[1] is formed from the circle diagram shown in Fig. 102 by a transformation of coordinates. The rectangular coordinates $r = R/Z_0$ and $x = X/Z_0$ are transformed into circular coordinates, two sets of orthogonal circles tangent at a point, as shown in the figure. The new rectangular coordinates are

$$u = \frac{r^2 + x^2 - 1}{(r + 1)^2 + x^2} \tag{175}$$

and

$$v = \frac{2x}{(r + 1)^2 + x^2} \tag{176}$$

Radiating from the center of the diagram ($u = 0$, $v = 0$) is the new radial coordinate equal to the ratio of the direct and reflected amplitudes

$$w = \sqrt{u^2 + v^2} = \frac{I_2}{I_1} \tag{177}$$

The angular coordinate about the center is our old friend

$$\beta l = \tan^{-1} \frac{u}{v} \tag{178}$$

[1] See reference, page 150.

As in the other forms of chart, distance in electrical radians away from the termination is measured clockwise.

The advantages of introducing these transformations are (1) all possible values of r and x are contained within the circle for $r = 0$, (2) the angular coordinate βl is uniformly distributed

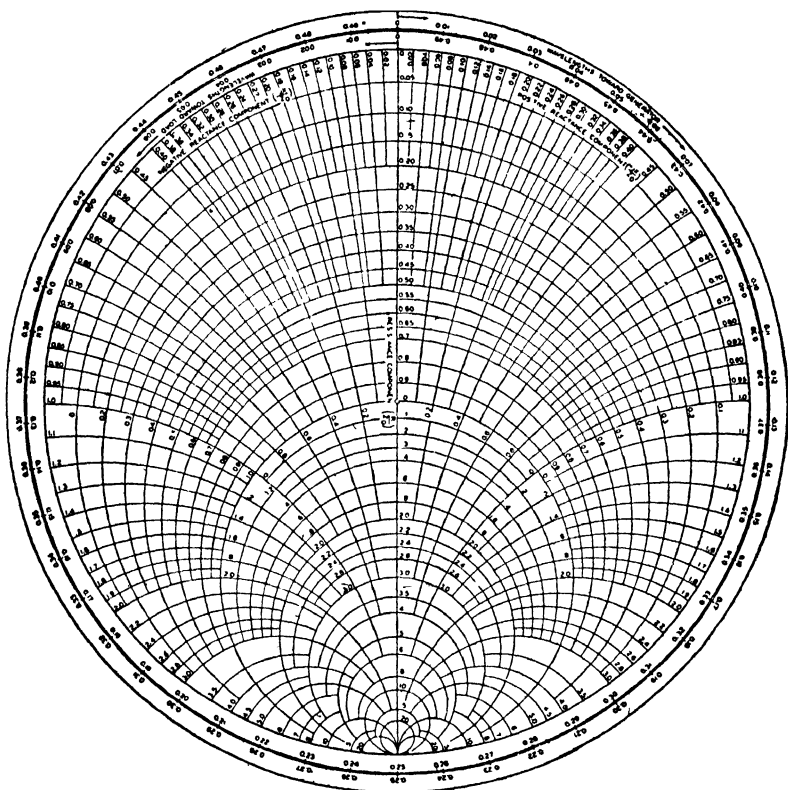



FIG. 107.—The Smith transformation of the circle diagram, widely used in practice.

about the $r = 0$ circle, and (3) the radial quantity w is equal to the ratio I_1/I_2 of the direct and reflected wave amplitudes at the termination of the line. The only disadvantage is that large values of r and x are difficult to locate accurately on the chart because of the small scale. Fortunately such large values are not often encountered in practice.

The polar coordinates w and βl are usually omitted from the chart to avoid confusion among lines In their stead a

radial ruler is employed, pivoted at the center of the diagram and marked with a nonlinear scale of the values of the standing-wave ratio,

$$k = \frac{I_1 + I_2}{I_1 - I_2} \tag{179}$$

where I_1 and I_2 are, respectively, the direct-wave amplitude and the reflected wave amplitude at the termination. The significance of the standing-wave ratio, which has not been discussed previously, will be clear in the next section, which treats power transfer and measurements.

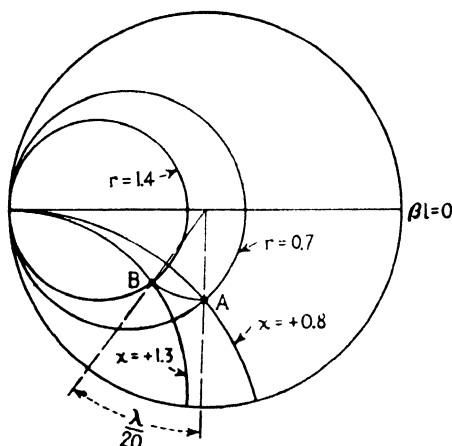



FIG. 108.—Construction on the Smith diagram to find change of impedance in going through 1/20 wavelength of line.

Two examples of the use of the Smith circle chart are shown in Figs. 108 and 109. In Fig. 108, a particular value of the normalized terminating impedance $z_t = r_t + jx_t$ is located at the point A. The line impedance at a distance βl electrical radians away from the termination is found by swinging an arc clockwise from point A to point B. The r and x coordinates at point B give the input impedance to the line at that point.

The radial distance of the point B from the center of the chart shows that the standing-wave ratio k is 2.3 ($w = 0.4$). The electrical distance from point A to $\beta l = 0$ (92 deg) shows the distance from the termination to the first voltage minimum in the standing-wave pattern. Figure 109 shows the inverse process, by which an unknown terminating resistance may be found from the standing-wave ratio and the position of the voltage minimum.

The same chart serves for admittance computations, with $g = G/Y_0$ and $b = B/Y_0$ substituted for r and x. Electrical distances measured from the termination are measured clockwise and inductive susceptance is found in the regions marked $-x$, conforming with the discussion of Sec. 69.

FIG. 109.—Construction on the Smith diagram to find the terminating resistance from nodal positions.

FIG. 110.—Introduction of series reactance to cancel reactive component, represented by rotation to center of chart.

The introduction of series reactance (or shunt susceptance) to the line is represented in the Smith chart as shown in Fig. 110. From a given point, corresponding to the position the reactance is inserted, the reactance is represented by proceeding along a

line of constant r, crossing the x lines by an amount equal to the magnitude of the inserted (normalized) reactance. The end of this line represents the impedance of the line after the reactance has been added. Figure 110 shows the removal of reflections by displacement to the point $r = 1$, $x = 0$ (the center of the chart). The case corresponds to Fig. 105.

71. Power Transmission and Measurements in Transmission Lines.—We have already stated that maximum power transmission is achieved when the reflected wave is minimized by matching impedances at each junction. We now examine this concept in more detail and show how the reflected waves are measured so as to permit adjustment for maximum power transfer.

To obtain a qualitative physical picture of the losses in transmission lines, let us consider a length of loss-free transmission line between an r-f generator at the input end and an r-f load at the output end, as shown in Fig. 111. The generator is matched to the characteristic impedance of the line, but the load impedance is not. Hence when the transmitted wave arrives at the load a part of it is reflected. The reflected wave travels back to the generator at the input end where, because the impedances are matched, it suffers no reflection. Instead it enters the resistive impedance in the generator and is wholly absorbed. Thus, in this case, no part of the reflected wave is retransmitted. It represents a loss of power that is expended simply in heating the generator.

Suppose now that the generator impedance is not matched to the characteristic impedance of the line. Then when the reflected wave arrives at the generator it is partly reflected back toward the load. This secondary reflected wave arriving at the load is again partly absorbed and partly reflected. As the wave travels back and forth between load and generator, at each reflection part of it is delivered to the load and part to the generator. In this way the energy of the primary reflected wave is distributed between load and generator until its amplitude becomes negligibly small.

If we now consider that the line has losses, during each traverse of the line, the reflected wave gives up part of its energy to the line itself and the amount of energy delivered to the load (and to the generator) is proportionately reduced. In any event, it is evident that the root of the trouble is the fact that the reflected

wave arose at the output end in the first place. If the load had been matched to the characteristic impedance, all of the power delivered to the load by the direct wave would have been absorbed by the load and the power losses would have been limited to those of *one* traverse of the line. If in addition the generator had been matched to the line impedance, the generator would have delivered a maximum amount of power to the line, and the amount of power in turn delivered to the load would have been proportionately greater. It is thus evident that a transmission

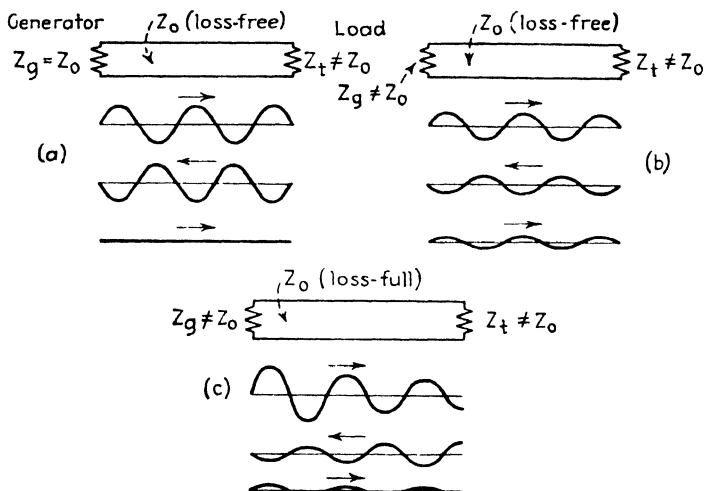



FIG. 111.—Reflection in transmission lines: (*a*) generator matched to loss-free line; (*b*) generator and load mismatched, loss-free; (*c*) generator and load mismatched, loss-full line.

lines hould be matched at *both* ends to effect maximum power transfer. It must be matched at the output end to minimize power losses in the line.

One further qualitative observation should be made. When reflected waves exist in the line, the vector addition of the direct and reflected waves results in a standing wave. The standing wave is evidenced by the existence of high voltage across the line at intervals separated by a half wavelength. When the power level is high, as during the transmitting period of radar operation, the regions of high voltage may add materially to the losses through the effects of corona and may in fact cause complete breakdown of the insulation. Thus standing waves

must be minimized to assure that the insulation is not stressed beyond safe limits.

The quantitative statement of power transfer is derived in terms of Fig. 112, which shows a generator of internal voltage E and internal impedance $Z_g = R_g + jX_g$ connected to a load of impedance $Z_t = R_t + jX_t$. The power P_t delivered to the load is the magnitude of the current squared times R_t.

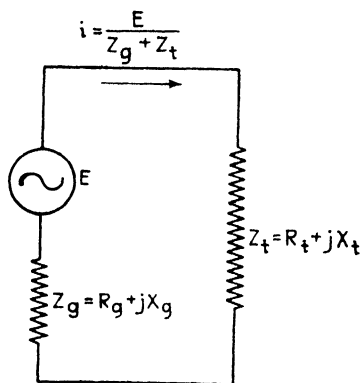



FIG. 112.—Power transfer from reactive-resistive generator to reactive-resistive load

$$P_t = |I|^2R_t = \frac{E^2R_t}{(R_g + R_t)^2 + (X_g + X_t)^2} \tag{180}$$

Inspection of this equation shows that P_t is a maximum when $R_g = R_t$ and $X_g = -X_t$. There are two cases of special interest: (1) when the reactances are opposite and equal but the resistances are not equal, and (2), the resistances are equal but the reactances are not balanced. In case (1)

$$P_t = \frac{E^2R_t}{(R_g + R_t)^2} \tag{181}$$

In case (2)

$$P_t = \frac{E^2/R_t}{4 + (X_g + X_t)^2/R_t^2} \tag{182}$$

These expressions are plotted in Fig. 113. It will be noted that the amount of power delivered falls off slowly with resistance mismatch and somewhat more rapidly with reactance mismatch.

Equation (180) may be used to determine the energy delivered by a transmission line to a load by calculation or measurement of the resistance and reactance present at the output termination. This power may then be compared with that delivered to the line at the input and the losses in the line thereby determined. But this is a cumbersome experimental procedure, which becomes progressively more difficult as the frequency of operation increases. An alternative procedure, measurement of standing-wave ratios, is relatively simple to perform at uhf and shf frequencies.

The standing wave formed by addition of the direct and reflected waves is shown in Fig. 114. When no reflections exist, the rms voltage amplitude along the line remains substantially constant, except for the slow decrease due to attenuation. When reflections are present, the rms amplitude varies periodically,

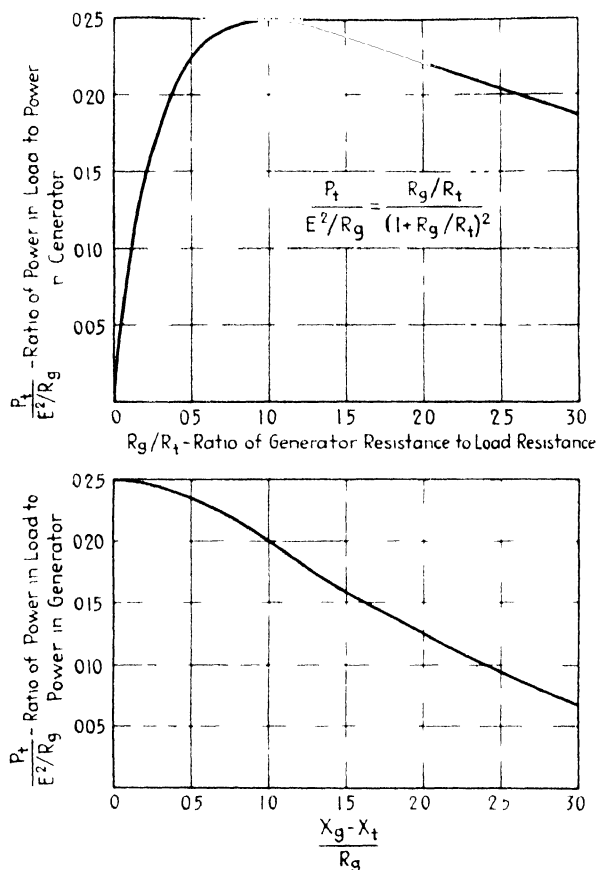



FIG. 113.—Power transferred from generator to load as a function of resistance mismatch (top) and reactance mismatch

displaying a maximum voltage V_{max} at certain points on the line, separated a half wavelength. A minimum voltage V_{min} appears at intermediate points, that is, one-quarter wavelength from the maxima. The measurement of these voltage maxima and minima reveals the wavelength within the line. It also indicates the amount of reflected power relative to that transmitted and

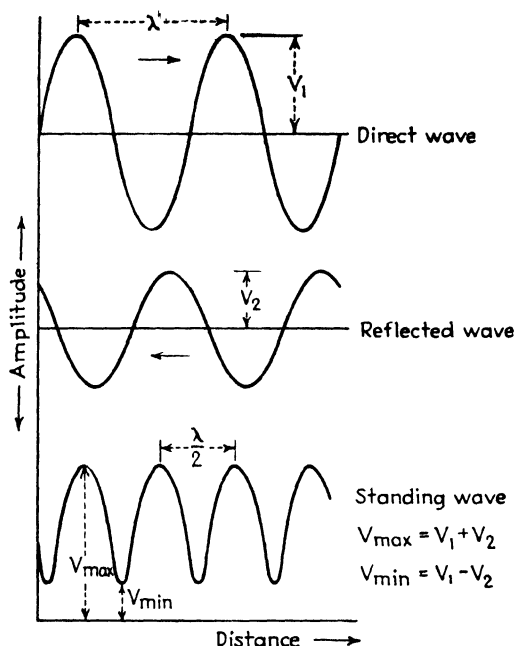



FIG. 114.—Formation of standing wave from direct and reflected waves.

absorbed by the load. To show this, we note from Fig. 114 that the voltage maximum arises from the sum of the direct and reflected components

$$V_{max} = V_1 + V_2 \tag{183}$$

whereas the voltage minimum arises from their difference

$$V_{min} = V_1 - V_2 \tag{184}$$

Hence the standing-wave voltage ratio is

$$k = \frac{V_{max}}{V_{min}} = \frac{V_1 + V_2}{V_1 - V_2} \tag{185}$$

By manipulation of Eq. (185) we obtain

$$\frac{V_2}{V_1} = \frac{k - 1}{k + 1} \tag{186}$$

But the power absorbed in the load is $P_1 = V_1^2/R_t$ and the power reflected is $P_2 = V_2^2/R_t$ (the resistive component of the load is

common to both voltages). Hence

$$\frac{P_2}{P_1} = \frac{V_2^2}{V_1^2} = \frac{(k-1)^2}{(k+1)^2} \tag{187}$$

The ratio of reflected to transmitted energy is thus directly obtained from the standing-wave voltage ratio k. Figure 115 indicates the percentage loss in reflected energy in terms of k.

The methods of measurement of k (see Chap. XII) by means of a thermocouple meter or bolometer indicate the quantity k^2,

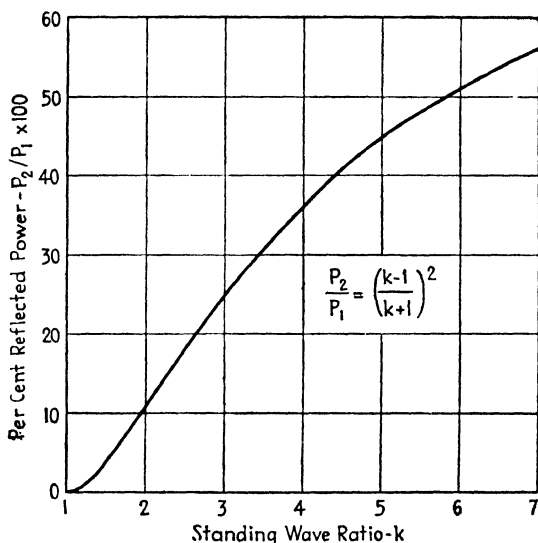



FIG. 115.—Power loss associated with standing wave ratio.

and the square root must be taken before entering the chart. It will be noted that substantial standing-wave voltage ratios are permissible (up to 2) before the loss in power exceeds 10 per cent or 0.5 db. This indicates that the standing-wave measurement is a sensitive measurement of power loss.

The standing-wave voltage ratio, in conjunction with the Smith circle diagram, provides a means of measuring the impedance at the end of a line. The method is as follows: The load end of the line is short-circuited and the position of a voltage minimum determined. The load impedance is connected and the shift in position of the voltage minimum away from the load

is noted. This gives a value of βl, in electrical radians, which may be laid out on the circle diagram (Fig. 116). The value of k is then measured out from the center of the diagram along the radius corresponding to the value of βl just determined. The point thereby established gives the r and x components of the load impedance. The purpose of short-circuiting the load

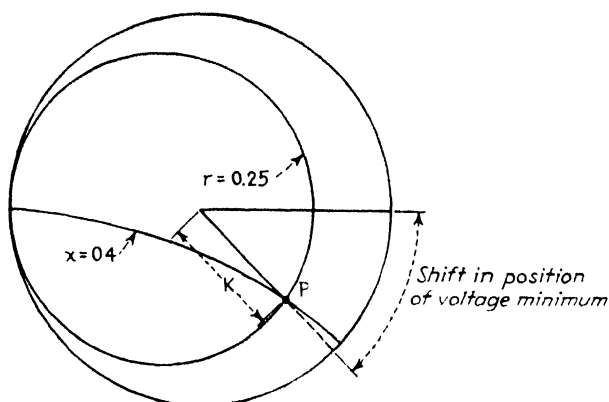



FIG. 116. Measurement of terminating impedance from standing-wave ratio and nodal positions.

is to establish a voltage minimum at the termination point, from which the voltage minimum under load conditions can be measured.

72. Parameters of Transmission Lines in Terms of Dimensions and Materials.—The foregoing discussion has indicated that the parameters (α, β, and Z_0), on which transmission line computations are based, are expressed in terms of the resistance R, conductance G, inductance L, and capacitance C per meter loop length of line. In this section we state, without derivation, the expressions that give these quantities in terms of the dimensions of the line and the properties of the materials used in its construction.

FIG. 117.—Parallel-wire transmission line.

Balanced Open-wire (Parallel-wire) Lines.—The dimensions of a balanced open-wire line are shown in Fig. 117. The parameters, in terms of these dimensions are

$$L = 0.921 \times 10^{-6} \log_{10} \frac{S}{r} \quad \text{henries per m} \tag{188}$$

$$C = \frac{0.120 \times 10^{-10}}{\log_{10} S/r} \quad \text{farads per m} \tag{189}$$

$$Z_0 = \sqrt{\frac{L}{C}} = 276 \log_{10} \frac{S}{r} \quad \text{ohms} \tag{190}$$

where S is the separation of the wires between centers and r is the radius of each wire (S and r in the same units). These expressions apply to lines immersed in air, which is the universal practice with open-wire lines.

An approximate expression for the loop series resistance of a copper line is

$$R = 8.4 \times 10^{-6} \frac{\sqrt{f}}{r} \quad \text{ohms per m} \tag{191}$$

where f is the frequency of operation in cycles and r is the wire radius in centimeters.

FIG. 118 Coaxial transmission line.

Coaxial (Concentric) Lines.—The dimensions of a coaxial line are shown in Fig. 118. The line parameters, in terms of these dimensions, are

$$L = 0.461 \times 10^{-6} \mu \log_{10} \frac{r_o}{r_i} \quad \text{henries per m} \tag{192}$$

$$C = \frac{0.241 \times 10^{-10} \epsilon}{\log_{10} r_o/r_i} \quad \text{farads per m} \tag{193}$$

$$Z_0 = \sqrt{\frac{L}{C}} = 138 \sqrt{\frac{\mu}{\epsilon}} \log_{10} \frac{r_o}{r_i} \quad \text{ohms} \tag{194}$$

where μ is the permeability (usually unity) and ϵ the dielectric constant (close to unity except in solid dielectric cables) of the space or material between the conductors, r_o is the inner radius of the outer conductor, and r_i the radius of the inner conductor (r_o and r_i in the same units).

Attenuation in Coaxial Lines.—The wide use of coaxial lines in radar and related radio practice has required that particular attention be paid to the attenuation in such lines.

The most important loss is usually caused by R, the series r-f resistance of the conductors. The shunt losses due to ohmic conductance are usually so low that they may safely be neglected. On the other hand, the shunt losses due to imperfections of the dielectric in the insulation are often appreciable. In solid dielectric cables at the higher frequencies, in fact, these dielectric losses are predominant.

The series losses depend on the resistivity of the inner and outer conductors and the skin depth through which the current flows. The skin depth is defined as the depth at which the current density falls to $1/2.718 = 1/\epsilon$ of the value at the surface. It is given by

$$\delta = \sqrt{\frac{\rho}{2\pi\omega\mu}} \qquad \text{cm} \tag{195}$$

where ρ is the resistivity of the conductor in abohm-centimeters (ohm-cm $\times 10^9$), $\omega = 2\pi f$ is the angular frequency of the current, and μ is the permeability of the conductor (unity in all conductors used in practice). At microwave frequencies in good conductors (copper, silver, brass, etc.) the skin depth is about 10^{-4} cm.

The resistance per unit length of a coaxial cable (inner and outer conductors made of the same material) is

$$R = \frac{100\rho}{2\pi\delta}\left(\frac{1}{r_0} + \frac{1}{r_i}\right) \qquad \text{ohms per meter} \tag{196}$$

where ρ is the resistivity of the conductors in ohm-centimeters, δ the skin depth in centimeters, r_0 and r_i in centimeters.

The attenuation is expressed in terms of the attenuation constant α in nepers per meter (8.69α = decibels per meter). Reference to Eq. (143) shows that when the conductance losses are very small, α is given by

$$\alpha = \frac{R\sqrt{C/L}}{2} \qquad \text{nepers per m} \tag{197}$$

Substituting Eqs. (194) and (196) in Eq. (197) produces

$$\alpha = \frac{100\rho(1/r_o + 1/r_i)}{2\pi\delta 276\sqrt{\mu/\epsilon}\log_{10} r_o/r_i} \qquad \text{nepers per m} \tag{198}$$

The presence of the skin depth in the denominator shows that

the attenuation tends to increase with the square root of the frequency. The optimum ratio of r_0/r_1 for minimum attenuation, indicated by Eq. (198), is about 3.6 but the value is not critical. The corresponding value of characteristic impedance is about 75 ohms.

Another attenuation term must be added to that of Eq. (198) to account for losses in the dielectric, particularly when solid dielectric is used. This term is given by

$$\alpha_d = \frac{\pi \sqrt{\epsilon_r} \tan (\pi/2 - \theta)}{\lambda} \quad \text{nepers per m} \qquad (199)$$

where ϵ_r is the real part of the complex dielectric constant, θ is the power factor angle of the dielectric, and λ is the operating wavelength in meters. Since this attenuation term increases with the *first power* of the frequency, attenuation due to dielectric losses tends to mask attenuation due to skin effect (which increases with the square root) at the higher frequencies.

Optimum Dimensions for Power Transfer.—The power delivered by a transmission line to a matched load is

$$P = \frac{V^2}{Z_0} \quad \text{watts} \qquad (200)$$

where V is the rms amplitude of the voltage at the termination and Z_0 is the characteristic impedance (equal to the termination resistance). The upper limit of power transferred is defined by the voltage at which the insulation of the cable breaks down. To permit the highest possible voltage, the voltage gradient within the cable should be kept as low as possible by proper choice of the dimensions. In a coaxial cable the gradient is lowest when $r_0/r_i = \epsilon = 2.718$, which corresponds to a characteristic impedance of 60 ohms. However, the maximum power will be transferred if the characteristic and load impedances are 30 ohms, corresponding to $r_0/r_i = 1.65$. Among the several preferred values of Z_0 (30 ohms for maximum power transfer, 60 ohms for maximum voltage, and 75 ohms for minimum attenuation) there is no sharp demarcation. Values from 50 to 100 ohms are customarily used.

73. Waveguides. General Considerations.—The waveguide[1] is essentially a hollow tube of conducting material that acts as a

[1] See bibliography at end of chapter.

boundary for electromagnetic waves and conducts them in much the same manner as a transmission line. The advantages of the waveguide relative to a coaxial line of the same cross-sectional area are (1) lower attenuation, (2) higher power-handling capacity, and (3) mechanical simplicity. The disadvantage is the fact that waveguides of convenient size are suitable for transmission of high frequencies only, in the thousands of megacycles. This limitation arises from the fact that guided waves are propagated only if the wavelength is smaller (in rectangular guide) than twice the maximum linear dimension of the cross section of the guide. When the wavelength is longer than this, the energy is very rapidly attenuated. The waveguide thus acts as a high-pass filter, passing frequencies above a certain limit determined by the size of the guide and attenuating all lower frequencies. A well-designed waveguide, properly adjusted, displays low attenuation over a wide band of frequencies above the critical limit.

The lower attenuation of the guide for a given cross section is readily explained in terms of conductor losses. Equation (198) shows that the attenuation of a coaxial line varies inversely with the radii of the conductors. The attenuation associated with the inner conductor thus predominates. In the waveguide, in effect, the inner conductor is removed and with it the associated attenuation. Moreover, the absence of insulation and stubs generally allows greater voltage gradients to exist within the guide than would be possible in a coaxial line of the same size. This fact accounts for greater power-handling ability.

At the very highest frequencies (10,000 megacycles and above) another comparison is of interest. To avoid propagation at higher modes in coaxial cable, the average perimeter of the cable cross section must be small compared with the wavelength. Thus at these very high frequencies, the coaxial line has a very small cross section and a correspondingly limited power capacity. The waveguide, at the same frequencies, can avoid propagation at higher modes so long as its maximum cross section dimension is less than one-half wavelength. This limit allows about twice as much cross-sectional area in the waveguide as is permissible in the corresponding coaxial line, and the attenuation and power capacity are thus both considerably in favor of the waveguide.

One indirect disadvantage of the waveguide is difficulty in understanding the mechanism by which the energy is propagated. This conceptual difficulty, encountered by those trained to deal only with circuit concepts, limits the application of waveguides to simple cases. The necessity of a more basic understanding, forced upon all concerned by the pressure of war activity, has led to a much wider appreciation of the capabilities and limitations of waveguides. This experience has proved that the application of guides to new applications, and even in stereotyped cases, demands that the technical worker have a basic understanding of the mechanism of wave propagation.

In what follows we shall present first a treatment similar to that of Skilling,[1] to indicate the physical nature of the propagation and then introduce a fairly complete derivation of the transverse electric (*TE*) modes in rectangular guide.

74. Mechanism of Propagation in Waveguides.—The manner in which energy is propagated along a waveguide is best introduced by considering a plane wave of electromagnetic radiation and endeavoring to fit it into the space within the guide. When we do so we find that physical conditions, which must be met at the walls of the tube, prohibit the transmission of a simple plane wave. Rather, a more complicated configuration of the electric and magnetic fields is required. In a particular case of great practical importance ($TE_{1,0}$ mode used in rectangular guides) the required field configuration may be considered as the sum of two plane waves that are propagated in zigzag fashion, by reflection, between opposite faces of the guide. This zigzag propagation explains the existence of the cutoff frequency and of such related phenomena as the low velocity of energy propagation and the long wavelength within the guide when the frequency is near the cutoff value.

Consider first the transverse, plane-polarized wave of Fig. 119*a*. The electric vector E and the magnetic vector H are vibrating sinusoidally at right angles to each other. The direction of energy propagation is indicated by the Poynting vector that is at right angles to both E and H. This is the familiar electromagnetic wave in empty space (that which exists, for example, at a great distance from a radiating antenna). Now suppose

[1] SKILLING, H. H., "Fundamentals of Electric Waves," Chap. XIII, John Wiley & Sons, Inc., New York, 1942.

that, as in Figure 119*b*, two parallel planes of perfectly conducting material are introduced in the field such that the planes are perpendicular to the electric vector E. These two planes will form the top and bottom of the waveguide in the subsequent discussion. We now inquire whether the plane wave can exist, without changing its configuration, between these plates. The answer is that the plane wave can so exist. This mode of propagation is, in fact, the *principal mode* previously noted as occurring in coaxial lines. To prove that the field can exist, we recall that electric lines of force can terminate only on them-

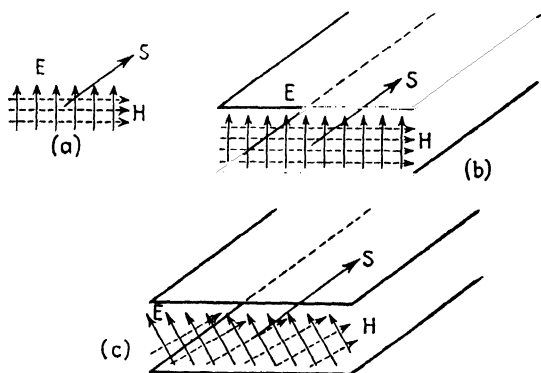



FIG. 119.—Plane electromagnetic wave (a) and two possible orientations with respect to parallel plates (b and c).

selves or on electric charges. If we assume that the plates are perfect conductors, charges will accumulate on the plates, under the lines of force, and thus terminate them. As the electric wave propagates between the conductors, the electric charges terminating the lines move also and thus create current in the conductors. The current thus formed must be identical to the current required by the motion of the magnetic field as it propagates between the conductors. Vector computation proves that both fields are satisfied by the same current, and thus the plane wave remains plane.

But this is true only so long as the electric vector is perpendicular to the conductors and the magnetic field is tangential to them. If the wave were oriented in any other fashion, as in Fig. 119*c*, there would be a component of electric field parallel to the surface, and a component of magnetic field perpendicular to the surface. These components would induce currents which

in turn would create electric and magnetic components exactly opposed in magnitude and direction to the first components. Thus only perpendicular electric field and tangential magnetic field can exist at the conductor surfaces. Evidently then, in Fig. 119c, the propagated wave would not be a simple plane wave, but would be distorted by virtue of the oblique polarization of the wave with respect to the conductors.

Now let us return to the initial (E-perpendicular, H-tangential) orientation and let us complete the waveguide by adding two side walls, as shown in Fig. 120a. Here we come upon a fundamental difficulty. The electric vector cannot be perpendicular to the side walls as well as to the top and bottom walls. Nor can the magnetic field be tangential to the side walls as well as to the top and bottom walls. But the requirements for electric perpendicularity and magnetic tangentiality remain in each case. Thus if we attempt to force a plane wave through a waveguide, even with the electric vector perpendicular to the top and bottom surfaces, the plane wave is inevitably distorted. In the particular case considered, the electric field E must vanish altogether at the side walls. It builds up, in sinusoidal distribution, to a maximum at the center of the guide. The magnetic field, to remain tangential to the side walls, turns through a right angle near the wall and forms a closed loop, coming back upon itself along the opposite side wall. These field distributions are shown in Fig. 120b. Since the electric field is purely transverse (no longitudinal component), this type of propagation is known as "transverse electric" (TE).

FIG. 120. - Waveguide formed by addition of sidewalls (a), which distort plane electromagnetic wave to transverse electric mode (b).

It is evident that the introduction of the side walls has distorted the plane wave into another form of wave that has no simple name. Fortunately, this distorted field can be represented as the sum of two plane waves, each traveling at an angle to the

side walls and hence reflected back and forth between them. Figure 121 shows that two such waves, properly disposed with respect to each other, can cancel each other at the side walls, as required in the resultant field shown in Fig. 120*b*. The same orientation of plane waves produces the required closed loop

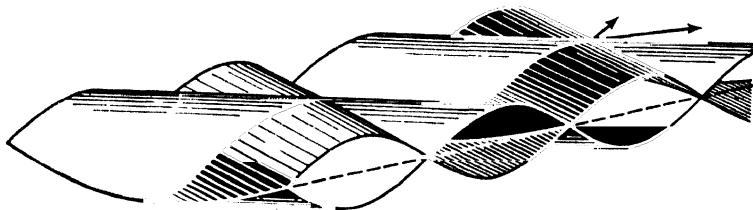



FIG. 121.—Three-dimensional representation of two component plane waves forming a transverse electric mode. (*After Skilling.*)

magnetic field. Thus we can omit further consideration of the resultant distorted field and concentrate on two superimposed plane waves.

The superimposed plane waves of Fig. 121 are propagated at an angle with respect to the side walls. We can ascertain

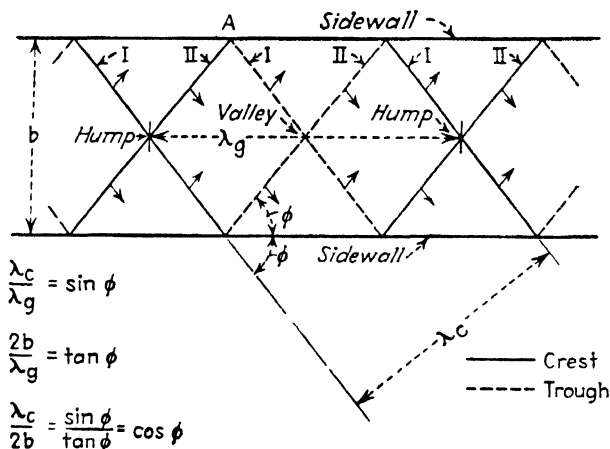



FIG. 122.—Representation of transverse electric mode as the sum of two plane waves propagated in zigzag fashion.

the value of the angle ϕ required to assure cancellation of the electric field along the side walls, by the following reasoning: Let the separation between the side walls be b cms as shown in Fig. 122. Suppose that the crest of one of the zigzag waves, at a particular instant of time, is at point A on the side wall.

Then the trough of the other wave must be at the same point at the same instant, since the sum of the two waves must be zero. The same balance must occur at the next crest and trough, respectively. Thus we have the combination of crests and troughs shown in Fig. 122. The perpendicular distance between crests is the wavelength of each zigzag wave. Inspection of the figure shows that this wavelength λ_c and the width of the guide b determine the angle ϕ by the following relation

$$\cos \phi = \frac{\lambda_c}{2b} \tag{201}$$

We note that the wider the guide becomes, relative to the wavelength of the zigzag waves, the larger the angle becomes. In the limit where the waveguide is infinitely wider than a wavelength $\phi = 90°$ and both waves are propagated directly down the guide as plane waves. The other limit is reached when the width of the guide is a half wavelength, that is, $b = \lambda_c/2$. Then by Eq. (201) ϕ is zero deg and the zigzag waves bounce back and forth perpendicularly to the guide walls. At this limit all the energy travels back and forth across the guide and none of it travels along the guide. This is the cutoff condition. When the wavelength is larger than twice the distance between sidewalls, no energy is propagated along the guide. Shorter wavelengths will be propagated, as if by two zigzag plane waves in superposition, the angle between wave and side wall being determined by the wavelength and the guide width, in accordance with Eq. (201).

75. Group Velocity, Phase Velocity, and Wavelength in Guide. The simplifying concept of the zigzag waves permits us to explain several important aspects of waveguide performance. In the first place the energy in each zigzag wave is propagated, along the zigzag path, at the speed of light, c cms per sec (assuming perfectly conducting walls).

The component of this velocity in the direction of the guide axis is determined by the angle ϕ and is equal to

$$v_g = c \sin \phi \tag{202}$$

But by Eq. (201)

$$\sin \phi = \sqrt{1 - \cos^2 \phi} = \sqrt{1 - \left(\frac{\lambda_c}{2b}\right)^2} \tag{203}$$

Thus the speed at which the energy propagates down the guide is

$$v_g = c\sqrt{1 - \left(\frac{\lambda_c}{2b}\right)^2} \tag{204}$$

This is the group velocity, so called because it is the velocity at which the group of sinusoidal components in an r-f pulse or other modulated signal are propagated down the waveguide. We note that the group velocity approaches zero as the wave-

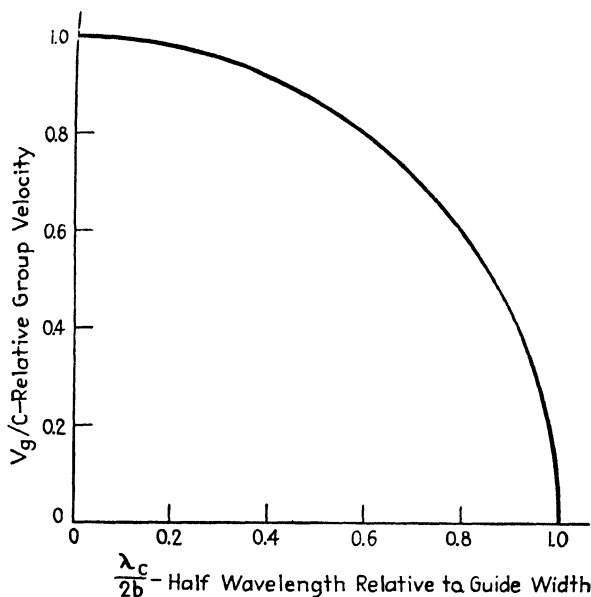



FIG. 123.—Variation of group velocity with wavelength relative to width of waveguide.

length approaches $2b$, twice the distance between side walls. This loss of signal velocity can be reduced by employing a guide whose width is greater than a half wavelength. The relationship of Eq. (204) is shown in Fig. 123.

There is another velocity in waveguide propagation, known as the "phase velocity." This is the velocity along the guide axis, of a surface of constant phase on the *resultant* wave, produced by the summation of the two zigzag waves. This summation is shown graphically in Fig. 124. We note that a hump of the resultant wave is produced by the superposition of two crests in the zigzag waves, as shown in Fig. 121. The speed with

which the hump (or any other distinguishing feature of the resultant wave) moves down the guide is the phase velocity. We note that this velocity is in effect a virtual velocity, which could be observed if the field configuration were visible, but which does not represent the propagation of energy down the guide.

FIG. 124.—Hump and valley form of resultant electric field in waveguide. (*After Chu and Barrow.*)

To find the phase velocity we refer to Fig. 122 and note that while the zigzag waves are moving a distance d along their paths, the hump formed by their summation moves along the guide a distance $d/\sin\phi$. Thus the phase velocity v_p is, by Eq. (203),

$$v_p = \frac{c}{\sin\phi} = \frac{c}{\sqrt{1 - (\lambda_c/2b)^2}} \tag{205}$$

We note that the phase velocity and group velocity are reciprocally related and that

$$v_g v_p = c^2 \tag{206}$$

The final aspect revealed by the zigzag waves is the wavelength between successive humps in the resultant wave. The wavelength of each zigzag wave along its path is the same as it would have in empty space:

$$\lambda_c = \frac{c}{f} \tag{207}$$

where f is the frequency of the wave. Similarly the wavelength in the guide (the distance between successive humps of the resultant wave, Fig. 124) is

$$\lambda_g = \frac{v_p}{f} = \lambda_c \frac{v_p}{c} \tag{208}$$

The reference to Eq. (205) shows that λ_g, the wavelength in the guide, is

$$\lambda_g = \frac{\lambda_c}{\sqrt{1 - (\lambda_c/2b)^2}} \tag{209}$$

This equation states the wavelength in the guide is generally longer than the wavelength of the same frequency in empty space. The "wavelength in guide" λ_g becomes rapidly longer

as the free-space wavelength λ_c approaches the cutoff value $2b$. Equation (209) applies generally to all modes of propagation and to all shapes of waveguide if the cutoff wavelength applicable to the mode and shape is substituted for $2b$.

The wavelength in the guide can be measured readily by a movable probe, inserted through the top or bottom surface, to indicate the standing-wave pattern. Such measurements show that the wavelength does in fact increase as the cutoff value is approached, in accordance with Eq. (209).

It is thus possible to deduce many of the operating parameters of a rectangular waveguide, without benefit of computation, simply by assuming sinusoidal propagation of two superimposed plane waves. It must be realized, however, that these operating parameters describe a particular case. Other cases may occur. If the guide is large enough, for example, the electric vector may become zero not only at the side walls but also along a plane midway between them. The derivation of the field configuration then proceeds as in the case just described, as if two waveguides were placed side by side, with the common sidewall between them removed. Moreover the electric field need not continue from the bottom to the top of the waveguide without diminution. A division of the electric field may occur along planes parallel to the top and bottom. Such divisions may be multiplied without limit, provided the guide is large enough with respect to the free-space wavelength.

Still another complication enters if the electric vector is polarized parallel to the axis of the guide. Then the magnetic field is purely transverse (no longitudinal component) and the propagation is known as "transverse magnetic" (*TM*). An infinite variety of modes is possible in this case, as in the transverse electric (*TE*) case just considered. Finally, solutions must be found for waveguides of circular cross section. Fortunately only a few modes are of practical interest in each type of guide.

To examine the higher order modes in waveguides, it is unfortunately necessary to resort to solutions of vector differential equations. Many excellent references[1] to this subject are available so no attempt will be made to cover all cases. It is

[1] See bibliography at end of this chapter

deemed worth while, however, to deduce the general equations for the TE modes in rectangular guide, if only as a more rigorous demonstration of the physical concepts advanced in the preceding paragraphs.

76. Solution of Transverse Electric Modes in Rectangular Guides.—The Maxwell equations, applicable to empty space within a waveguide, expressed in the symbols of vector calculus, are

$$\nabla \cdot \mathbf{E} = 0 \tag{210}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{211}$$

$$\nabla \times \mathbf{H} = \frac{1}{c}\frac{\delta \mathbf{E}}{\delta t} \tag{212}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\delta \mathbf{H}}{\delta t} \tag{213}$$

where $\mathbf{E}$ is the electric intensity vector, $\mathbf{H}$ the magnetic intensity vector, c the velocity of light, and t the time. These equations state, respectively: Eq. (210) that no free charges exist in the space; Eq. (211) that no free magnetic poles exist; Eq. (212) that any change of the electric field with time produces a magnetic field, whose curl is as stated; Eq. (213) that any change in the magnetic field with time produces an electric field whose curl is as stated.

These equations can be combined to form two equations of the second degree, known as the wave equations. The process is as follows: take the curl of both sides of Eq. (213)

$$\nabla \times (\nabla \times \mathbf{E}) = -\frac{1}{c}\nabla \times \frac{\delta \mathbf{H}}{\delta t} = -\frac{1}{c}\frac{\delta}{\delta t}(\nabla \times \mathbf{H}) \tag{214}$$

Substituting Eq. (212), this becomes

$$\nabla \times (\nabla \times \mathbf{E}) = -\frac{1}{c^2}\frac{\delta^2 \mathbf{E}}{\delta t^2} \tag{215}$$

But $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E}$, and by Eq. (210) $\nabla \cdot \mathbf{E} = 0$. Hence

$$\nabla^2\mathbf{E} = \frac{1}{c^2}\frac{\delta^2 \mathbf{E}}{\delta t^2} \tag{216}$$

By an identical process

$$\nabla^2\mathbf{H} = \frac{1}{c^2}\frac{\delta^2 \mathbf{H}}{\delta t^2} \tag{217}$$

Equations (216) and (217) are the electromagnetic wave equations in empty space. The vector symbols of Eq. (216) state that the sum of the second partial derivatives of a particular component of the electric intensity with respect to the three coordinate axes is proportional to the second partial derivative of that component with respect to time. In expanded form, Eq. (216) is

$$\begin{aligned} \frac{\delta^2 E_x}{\delta x^2} + \frac{\delta^2 E_x}{\delta y^2} + \frac{\delta^2 E_x}{\delta z^2} &= \frac{1}{c^2}\frac{\delta^2 E_x}{\delta t^2} \\ \frac{\delta^2 E_y}{\delta x^2} + \frac{\delta^2 E_y}{\delta y^2} + \frac{\delta^2 E_y}{\delta z^2} &= \frac{1}{c^2}\frac{\delta^2 E_y}{\delta t^2} \\ \frac{\delta^2 E_z}{\delta x^2} + \frac{\delta^2 E_z}{\delta y^2} + \frac{\delta^2 E_z}{\delta z^2} &= \frac{1}{c^2}\frac{\delta^2 E_z}{\delta t^2} \end{aligned} \tag{218}$$

and similarly for Eq. (217).

The solution of Eqs. (218) consists in finding variations of E with respect to x, y, z, and t, which will satisfy the differential relationships stated. As is usual in solving differential equations, it is convenient to assume a form of solution, perform the indicated operations, and see if an identity results. First we assume that the variation of the electric field with time is sinusoidal (which we know from previous experience to be true), using the imaginary exponent form

$$\mathbf{E} = E_{x,y,z}\epsilon^{j\omega t} \tag{219}$$

where $E_{x,y,z}$ is the vector amplitude, as a function of the x, y, and z coordinates, and $\omega = 2\pi f$ is the angular frequency of the wave. Then

$$\frac{1}{c^2}\frac{\delta^2\mathbf{E}}{\delta t^2} = -\frac{\omega^2}{c^2}E_{x,y,z}\epsilon^{j\omega t} \tag{220}$$

and

$$\frac{\delta^2\mathbf{E}}{\delta x^2} = \frac{\delta^2 E_{x,y,z}}{\delta x^2}\epsilon^{j\omega t} \tag{221}$$

Substituting these and similar forms in Eq. (218) and thence back to the initial wave Eqs. (216) and (217) we obtain

$$\nabla^2 E_{x,y,z} + \frac{\omega^2}{c^2}E_{x,y,z} = 0 \tag{222}$$

and

$$\nabla^2 H_{x,y,z} + \frac{\omega^2}{c^2}H_{x,y,z} = 0 \tag{223}$$

By this manipulation we remove the time factor in the equations and concentrate on the space variations, $E_{x,y,z}$ and $H_{x,y,z}$. Here

again we note that a sinusoidal variation is appropriate, since the second derivative of a sine is proportional to a sine, and Eq. (222) states that the second derivative of $E_{x,y,z}$ must equal $E_{x,y,z}$ multiplied by $-\omega^2/c^2$.

Thus we find that an electromagnetic wave in free space may vary sinusoidally in time and sinusoidally along the coordinate axes of space. Then, by means of the Fourier synthesis discussed in Chap. II, we may build up any required variation of the electromagnetic field by combinations of sinusoidal components.

FIG. 125.—Basic geometry of rectangular waveguide.

The application of the spatial-wave Eqs. (222) and (223) to a rectangular waveguide is indicated in Fig. 125. The axis of the guide is taken parallel to the z axis, and the guide is a cms wide along the x axis and b cms wide along the y axis. We now inquire about the variation of $E_{x,y,z}$ and $H_{x,y,z}$ with x, y, and z. These variations are the field configurations or "modes" possible within the guide. In the present discussion we limit our consideration to transverse electric (TE) waves, that is, those which have no longitudinal component E_z along the axis of the guide.

We assume that the electric field varies cosinusoidally with time and along the z axis (in the direction of propagation along the guide). We inquire about the distributions of the electric field on the cross section of the guide, that is, the x component

$$E(x,z,t) = E_x \cos (\omega t - \beta z) \qquad (224)$$

and y component

$$E(y,z,t) = E_y \cos (\omega t - \beta z) \qquad (225)$$

Here ω is the angular frequency and $\beta = 2\pi/\lambda_g$ where λ_g is the wavelength along z axis. Since E_z and all its derivatives are zero, we can write, from Eqs. (222), (224), and (225)

$$\frac{\delta^2 E_x}{\delta x^2} + \frac{\delta^2 E_x}{\delta y^2} = -k^2 E_x \qquad (226)$$

$$\frac{\delta^2 E_y}{\delta x^2} + \frac{\delta^2 E_y}{\delta y^2} = -k^2 E_y \qquad (227)$$

where $k^2 = \omega^2/c^2 - \beta^2$.

Here again we note that the sum of second derivatives of a quantity equal the quantity times a constant, and we conclude that E_x and E_y vary sinusoidally or cosinusoidally with x and y.

What remains to be determined is the periodicity of these sinusoidal variations. The periodicity is determined by the condition that there shall be no tangential component of electric field at any surface of the guide. Thus at the two side walls, $y = 0$ and $y = b$, E_x must be zero, and at the top and bottom walls, $x = 0$ and $x = a$, E_y must be zero.

So long as these restrictions are observed, E_y and E_x may vary sinusoidally across the a and b dimensions through as many cycles as the excitation conditions may dictate. Generally, E_x and E_y may go through m half cycles between side walls and n half cycles between top and bottom. Moreover E_y must have a maximum value at $y = 0$ and $y = b$ and E_x must have a maximum value at $x = 0$ and $x = a$, with cosinusoidal variation between these limits. Finally, the electric lines of force may curve so that they are perpendicular to two walls at right angles, provided only that

$$\frac{\delta E_y}{\delta y} + \frac{\delta E_x}{\delta x} = 0 \tag{228}$$

which is the applicable form of Eq. (210). All these conditions (zero field at certain walls, finite field at others) are met if the field components are given by

$$E_y = -\frac{n}{a} K \cos \frac{m\pi y}{b} \sin \frac{n\pi x}{a} \sin (\omega t - \beta z) \tag{229}$$

and

$$E_x = +\frac{m}{b} K \sin \frac{m\pi y}{b} \cos \frac{n\pi x}{a} \sin (\omega t - \beta z) \tag{230}$$

where K is an arbitrary proportionality constant.

These equations show that for every value of m and n there is a corresponding field distribution. Each such configuration is designated as the $TE_{m\,n}$ mode. The integers m and n may have any positive value, including 0, up to infinity, and therefore the possible field configurations are infinite in number. The values of m and n met in practice are, fortunately, limited to 0, 1, and 2.

When n is zero, there is no sinusoidal variation of the field between top and bottom of the guide, and the field is constant

along that direction. Similarly there is no horizontal variation for $m = 0$. Since the field cannot be constant in both directions at once in a waveguide, the $TE_{0,0}$ or TEM mode (the principal mode) cannot exist physically.

By similar reasoning the components of the magnetic intensity vector in TE propagation are found to be

$$H_x = \frac{n}{a} L \cos \frac{m\pi y}{b} \sin \frac{n\pi x}{a} \sin (\omega t - \beta z) \tag{231}$$

$$H_y = \frac{m}{b} L \sin \frac{m\pi y}{b} \cos \frac{n\pi x}{a} \sin (\omega t - \beta z) \tag{232}$$

$$H_z = M \cos \frac{m\pi y}{b} \cos \frac{n\pi x}{a} \cos (\omega t - \beta z) \tag{233}$$

where $L = -Mk^2/(\beta\pi)$ and M are arbitrary proportionality constants.

The validity of Eqs. (229) through (233) may be tested by performing the operations indicated in Eqs. (216) and (217), by noting that an identity results, and by noting that the boundary conditions are satisfied.

The modes described by these E and H components are shown in graphical form in Fig. 126 for values of m and n up to 2. The mode discussed in Secs. 74 and 75 is now recognizable as the $TE_{1,0}$ mode. This is the mode used almost exclusively in rectangular guide in practical applications.

FIG. 126.—Electric and magnetic fields of various transverse electric modes in rectangular guide.

Solutions of the field components can be found by a similar process when E_z is finite and $H_z = 0$, that is, when there is no longitudinal component of magnetic field. The expressions are very similar to Eqs. (229) through (233) inclusive, except that E and H are interchanged and that other arbitrary constants of proportionality apply. These modes are known as

the "transverse magnetic modes" and are designated as $TM_{m,n}$ modes.

The cutoff wavelength in a rectangular waveguide propagating in mode m and n (either TE or TM) is given by

$$\lambda_0 = \frac{2}{\sqrt{(m/b)^2 + (n/a)^2}} \tag{234}$$

This is the wavelength as it would appear in free space. In the guide this wavelength is, of course, infinitely long. We note that Eq. (234) reduces to $\lambda_0 = 2b$, for $m = 1$, $n = 0$, agreeing with the discussion of Secs. 74 and 75.

The wavelength in the guide λ_g is related to the wavelength in free space λ_c and the cutoff wavelength in free space λ_0 by

$$\lambda_g = \frac{\lambda_c}{\sqrt{1 - (\lambda_c/\lambda_0)^2}} \tag{235}$$

Equation (209) will be recognized as a special case of Eq. (235) when $\lambda_0 = 2b$, the value for the $TE_{1,0}$ mode.

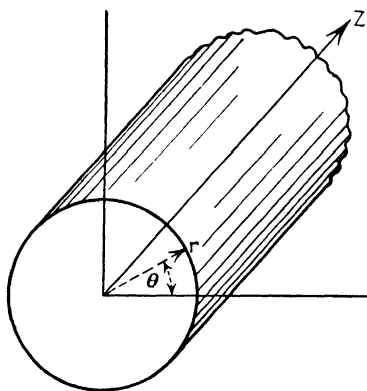



FIG. 127.—Basic geometry of circular waveguide.

77. Modes in Circular Waveguide.—Transverse electric (TE) and TM waves are propagated in guides of circular cross section by essentially the same physical action as in rectangular guides. To compute the geometry of the modes in circular guide, it is necessary to set up the Maxwell equations in cylindrical coordinates r, z, and θ, as in Fig. 127. The solutions to those equations are not sinusoidal, but take the form of Bessel functions, which are somewhat similar to the sine function with the amplitude of the harmonic variation decreasing along the radius of the guide cross section.

As in the case of the rectangular guide, we assume that the electric and magnetic fields vary sinusoidally with time at angular frequency ω and with distance z along the guide, the wavelength in the guide being $\lambda_g = \beta/2\pi$. We then inquire

what is the distribution of the field across the cross section of the guide, that is, the variation with r and θ. The solutions for the electric and magnetic fields in the TE modes are as follows:[1]

$$E_r = B\frac{\omega n}{k^2 r} J_n(kr) \sin n\theta \sin (\omega t - \beta z) \tag{236}$$

$$E_\theta = B\frac{\omega}{k} J'_n(kr) \cos n\theta \sin (\omega t - \beta z) \tag{237}$$

$$H_z = BJ_n(kr) \cos n\theta \cos (\omega t - \beta z) \tag{238}$$

$$H_r = B\frac{\beta}{k} J'_n(kr) \cos n\theta \sin (\omega t - \beta z) \tag{239}$$

$$H_\theta = -B\frac{\beta n}{k^2 r} J_n(kr) \sin n\theta \sin (\omega t - \beta z) \tag{240}$$

Here J_n is the Bessel function of order n, J'_n is the first derivative of J_n with respect to kr, and B is an arbitrary constant. We

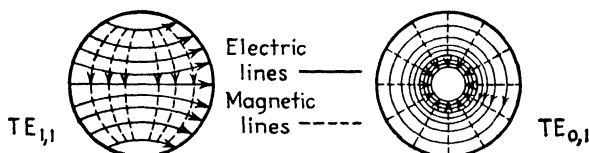



FIG. 128.—Transverse electric modes in circular guides.

note that the term sin $n\theta$ or cos $n\theta$ appears in each of these equations. Thus n is the number of complete cycles the configuration of the field completes about the center of the cross section of the guide. This is the angular type of field variation over the cross section. The other type of variation is radial, as is indicated by the presence of r in the argument of the Bessel functions. We recall that the tangential electric field E_θ must vanish, and the radial field E_r must have its maximum value at $r = a$. We note also that there may be any number m of maxima and minima in the field variation between $r = 0$ (the center of the pipe) and $r = a$ (the wall). Thus we find that the transverse electric modes $TE_{n,m}$ in circular guides are described by two subscripts, m the number of cyclic variations along the radius of the cross section and n the number of cyclic variations around the center of the cross section. As in the case of the rectangular

[1] For a detailed statement of the derivation, see J. G. Brainerd, and others, "Ultra-High-Frequency Techniques," p. 456, D. Van Nostrand Company, Inc., New York, 1942.

guide, only the lowest values of m and n, from 0 to 2, are of practical importance. Figure 128 shows the configurations of two of these lower modes. Note that the subscripts are written in the order n and m in circular modes.

The modes of transverse magnetic (*TM*) propagation are similar. The variation of the magnetic components H_r and H_θ across the cross section is given by expressions very similar to Eqs. (236) and (237), with E interchanged with H and with another arbitrary constant of proportionality Figure 129 shows two

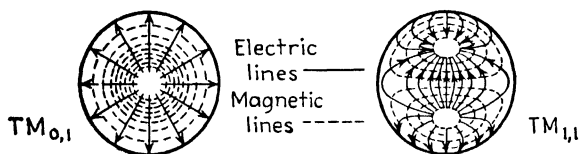



FIG. 129—Transverse magnetic modes in circular guides

of the lower *TM* modes in rectangular and circular waveguides.

The free-space cutoff wavelength in circular guides is equal to the radius of the guide times a constant, the constant being specified in terms of m and n. Thus

$$\lambda_0 = \left(\frac{2\pi}{x_{m\,n}}\right) a \qquad \text{cms} \tag{241}$$

where a is the radius of the guide cross section in centimeters and $x_{m,n}$ is the root of the Bessel function (or its first derivative) of order n that produces the mth zero. For *TE* modes $x_{m\,n}$ is the value of x in $J'_n(x) = 0$, and for *TM* modes it is the value of x in $J_n(x) = 0$. Values of $x_{m\,n}$ are given in Table VI.

TABLE VI.—VALUES OF $x_{m\,n}$

TE modes			*TM* modes		
	$m = 1$	$m = 2$		$m = 1$	$m = 2$
$n = 0$	3 832	7.016	$n = 0$	2 405	5 520
$n = 1$	1.842	5.330	$n = 1$	3 832	7 016
$n = 2$	3 050	6 710	$n = 2$	5 135	8 417

To obtain the wavelength in a circular guide at any frequency, we insert the value of λ_0 from Eq. (241) in Eq. (235), which applies regardless of the shape of the guide or mode of propaga-

tion. Knowing the wavelength in the guide, the phase velocity is computed from Eq. (208) and the group velocity from Eq. (206). In this manner several important practical aspects of circular waveguide operation can be deduced without specific reference to the field configurations.

78. Attenuation in Waveguides.—The field configurations have been computed in the previous sections on the assumption that the walls of the guide have perfect conductivity. Such an ideal waveguide would have no attenuation at frequencies above the cutoff frequency. But practical waveguides fall short of the ideal, since the walls must have finite conductivity. When such is the case, the components of the electric field perpendicular to the wall are no longer truly perpendicular but lean forward slightly in the direction of the guide axis. There is thus produced, directly at the wall of the guide, a tangential component of electric field E_t. This vector, acting with the tangential component H_t of magnetic field at the surface, forms a Poynting vector component that is directed into the wall. The Poynting vector represents a flow of energy into the side walls, that is, a loss of power abstracted from the guided wave and consumed in heating the wall.

In the presence of finite but high conductivity, the tangential magnetic component H_t gives rise to a small tangential electric component

$$E_t = H_t \sqrt{\frac{\mu\omega}{2\sigma}}\,(1 + j) \tag{242}$$

where μ is the permeability of the metal wall, ω is the angular frequency, σ is the conductivity, and j is the imaginary operator.

The Poynting vector component normal to the side wall is given by

$$\mathsf{S} = \frac{1}{2}\,Re(\mathbf{E} \times \bar{\mathbf{H}}) \tag{243}$$

where $\bar{H}$ is the complex conjugate of H. Substituting H_t and E_t in Eq. (242), we obtain

$$\mathsf{S} = \frac{1}{2}\sqrt{\frac{\mu\omega}{2\sigma}}\,|H|_t{}^2 \tag{244}$$

It is usual to express this loss of power as a fraction of the power

transmitted, in the form of an attenuation constant α. Thus defined

$$\alpha = \frac{1}{2} \frac{Re\int(\mathbf{E} \times \overline{\mathbf{H}})\, da}{Re\int(\mathbf{E} \times \overline{\mathbf{H}})\, da} \qquad \text{nepers per unit length} \qquad (245)$$

where the integration in the numerator is taken over the metal surfaces of the guide, through which energy is lost, and the integration in the denominator is taken over the cross section of the guide, across which the guided power is transmitted.

It is evident that the integrations depend on the size and shape of the guide as well as the mode of propagation. Hence the attenuation in rectangular guides differs slightly from that in circular guides of the same peripheral size, and in general the attenuation varies with the type of mode (TM or TE) and the values of the subscripts (m and n).

We shall indicate the manner of carrying out the integration for the $TE_{1,0}$ mode in rectangular guide. In the $TE_{1,0}$ mode the component fields are given in Eqs. (229) to (233). The values of the magnetic fields at the side walls are found by substituting $y = 0$, $y = b$, $x = 0$, and $x = a$ in order to find the tangential magnetic fields H_y and H_z at the side walls and H_x at the top and bottom walls. The Poynting vector given by Eq. (244) is then computed from these values of H_t, and each is averaged over a full period of the wave (over a wavelength in guide). The sum of these Poynting components is the total power loss to the guide walls per unit area. The Poynting vector representing the propagation of energy along the guide is then found from H_x, H_y, E_x, and E_y in Eqs. (229) to (233), and it is integrated over the cross section. The ratio of the two powers (one lost, and the other transmitted) is then taken, as in Eq. (245). The result, given here without proof[1] is

$$\alpha = \frac{2}{b}\sqrt{\frac{\epsilon_0}{\mu_0}}\sqrt{\frac{\mu\omega}{2\sigma}}\,\frac{(\lambda_0/2b)^2 + (b/2a)}{\sqrt{1 - (\lambda_0/2b)^2}} \qquad \text{nepers per cm} \qquad (246)$$

where a and b are the cross-sectional dimensions in centimeters, ω is the angular frequency, μ is the permeability, and σ is the conductivity of the guide walls ϵ_0 and μ_0, the dielectric constant

[1] The details of this derivation are given in J. C. Slater, "Microwave Transmission," Chap. III, pp. 138–150, McGraw-Hill Book Company, Inc., New York, 1942.

and permeability of the empty space within the guide, and λ_0 is the cutoff wavelength in centimeters. It is evident that the arbitrary proportionality constants K, L, and M of Eqs. (229) to (223) have disappeared in taking the ratio indicated in Eq. (245). The attenuation in decibels per centimeter is equal to 8.69α.

Practical values of attenuation constant may be found in handbooks.[1] The two attenuation formulas of most practical

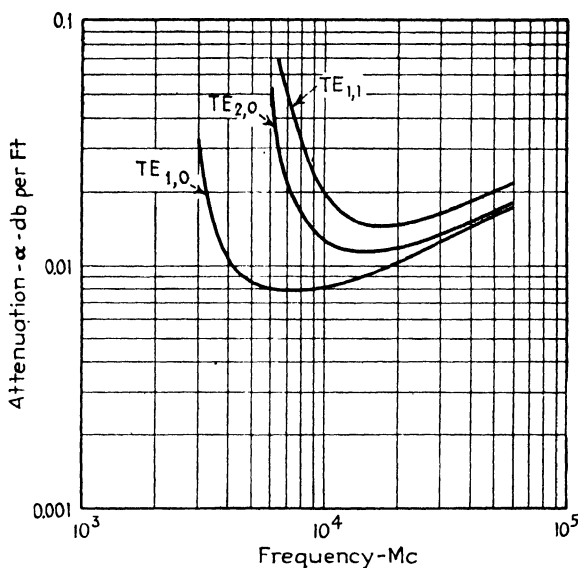



FIG. 130.—Attenuation in rectangular copper waveguide, inside dimensions 1 inch by 2 inches. (*After Moreno.*)

interest are those describing the loss in the $TE_{1,0}$ mode in rectangular guide and the $TE_{1,1}$ mode in circular guide. These are the *dominant modes* (those having the longest cutoff wavelengths) and are widely used in practice. The attenuation constants for these modes, in guides made of copper, are as follows:

$TE_{1,0}$ *rectangular*

$$\alpha = \frac{0.01107}{a^{3/2}} \left[\frac{\frac{1}{2}\, b/a\, (\lambda_0/\lambda)^{3/2} + (\lambda_0/\lambda)^{-1/2}}{\sqrt{(\lambda_0/\lambda)^2 - 1}} \right] \text{ db per ft.} \quad (247)$$

[1] See TERMAN, F. E., "Radio Engineers' Handbook," Sec. 3, pp. 260–264, McGraw-Hill Book Company, Inc., New York, 1943.

where λ is the operating free-space wavelength, λ_0 the cutoff free-space wavelength ($=2b$), both in centimeters, and a and b are the guide cross-sectional dimensions in centimeters. Figure 130 shows the attenuation in rectangular guides of this and other modes.

$TE_{1\,1}$ circular

$$\alpha = \frac{0.00423}{a^{3/2}}\left[\frac{0.420(\lambda_0/\lambda)^{3/2} + (\lambda_0/\lambda)^{-1/2}}{\sqrt{(\lambda_0/\lambda)^2 - 1}}\right] \quad \text{db per ft} \qquad (248)$$

where a is the radius of the guide cross section in centimeter λ the operating free-space wavelength, and λ_0 the cutoff free-

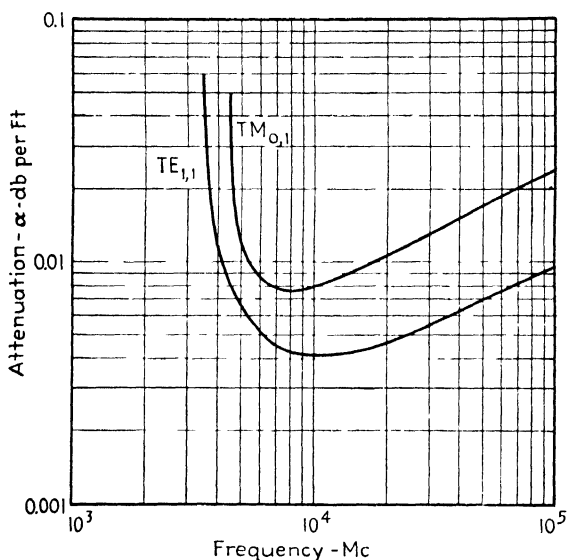



FIG. 131.—Attenuation in circular copper waveguide, inside diameter 2 inches. (*After Moreno.*)

space wavelength, both in centimeters. Figure 131 shows the attenuation of several modes in circular waveguides. It will be noted that the losses are roughly the same in rectangular and circular waveguides, when the cross-sectional periphery of each is the same.

The attenuation of frequencies below the cutoff value is of interest whenever such modes are generated (by discontinuities in the guide). The attenuation constant for modes below cutoff

is

$$\alpha = 8.69\sqrt{\left(\frac{2\pi}{\lambda_0}\right)^2 - \left(\frac{2\pi}{\lambda}\right)^2} \quad \text{db per cm} \tag{249}$$

where the free-space wavelength λ and the cutoff wavelength λ_0 are in centimeters. For very large operating wavelengths, Eq. (249) reduces to

$$\alpha = \frac{54.6}{\lambda_0} \quad \text{db per cm} \tag{250}$$

In a $TE_{1\,0}$ rectangular guide that has a cutoff wavelength of 11 cm (dimension b = 5.5 cm), the attenuation for low frequencies is about 5 db per cm, or about 150 db per ft (power loss ratio of 10^{-15} per ft). It is evident that l-f modes are attenuated so rapidly that they cannot be detected at all except in the near vicinity of the discontinuity causing them to appear.

79. Choice of Mode and Dimensions of Waveguides.—It would appear that the designer has a wide choice in selecting the mode of operation in a waveguide. By choosing the method of exciting the waveguide at its input end, it is possible to establish any one of a large number of modes. It might then be imagined that the choice between available modes should be made on the basis of minimum attenuation, of maximum power-handling ability, or of some other electrical property of economic importance. Unfortunately the choice cannot be made only on this basis. While a particular mode may be established initially in a guide, it is very difficult, if not impossible, to prevent the establishment of lower modes simultaneously. If the guide were infinitely long and perfectly uniform in cross section and in electrical constants, one mode might be sustained to the exclusion of others. But guides must be terminated in practice, and the termination generally involves standing waves which must be compensated by a reactive tuning device, the setting of which depends on the wavelength in the guide. When such compensation of standing waves is set up for a particular mode, the impedance match is correct for that mode only. Reflected waves occur for all other modes. Hence at each discontinuity or termination there is a mechanism for the transfer of energy from the desired modes to other modes. Since the other modes do not draw as much energy from the source as the

matched mode, the tendency is toward a large amplitude in the other modes.

To avoid this situation, *it is necessary to choose the dimensions of the guide and method of excitation such that only* one mode can be propagated, regardless of the discontinuities or terminations present. The restriction to one mode is accomplished by constructing the guide just large enough to contain one half cycle of the mode having the longest wavelength in guide (the dominant mode), but not large enough to allow the formation of two half cycles of the mode having the next smaller wavelength in guide.

In rectangular guide, with the electric field parallel to the Y axis, the dominant mode is $TE_{1,0}$. Inspection of Eqs. (234) and (235) shows that this choice of $m = 1$ and $n = 0$ corresponds to the longest wavelength in the guide for a particular frequency of excitation. Suppose λ is the free-space wavelength to be transmitted. Then the width of the guide should be greater than $b = \lambda/2$, in order to accommodate one half cycle of the dominant mode. But the guide must not be as wide as $b = \lambda$, because such a width would permit the transmission of two half cycles of the $TE_{2,0}$ mode. Thus there are lower and upper limits ($\lambda/2$ and λ) on the width of the guide.

The height of the guide is not critical, since the field does not vary in this direction. But too small a value is not desirable because of the tendency of the guide to sparkover, in the direction of the electric field, when the power level is high. Moreover the height cannot be greater than $a = \lambda/2$, since then a half cycle of vertical variation of electric field would be possible and mode $TE_{1,1}$ might be propagated.

Thus if b is λ or smaller (but not smaller than $\lambda/2$) and a is $\lambda/2$ or smaller, only the $TE_{1,0}$ mode can be propagated. The limiting cross sections are a square $\lambda/2$ on a side and a rectangle not larger than λ wide by $\lambda/2$ high. The rectangular form is more commonly used, because of its greater cross-sectional perimeter and hence lower attenuation. Generally, because standard rectangular pipe is thus manufactured, the width is twice the height in outside dimension. The ratio of the inside dimensions (which govern the propagation) is somewhat greater than 2:1. When such dimensions are chosen, the $TE_{1,0}$ mode has generally lower attenuation than the higher TE and TM modes. Attenua-

tions of from 0.005 db per ft to 0.05 db per ft are commonly achieved in copper guides, and lower attenuation is possible if the inside of the guide is silver-plated to lower the resistivity.

In circular guide, the dominant mode is $TE_{1,1}$ as can be seen from inspection of Eq. (241) and Table VI. The cutoff wavelength for this mode is $\lambda_0 = 3.4a$. Hence the radius of the guide cross section must be $a = \lambda/3.4$ or greater to allow transmission of a free-space wavelength λ. The next shorter cutoff wavelength is that of $TM_{0,1}$, for which $\lambda_0 = 2.6a$. Hence if the guide radius is smaller than $a = \lambda/2.6$, the $TM_{0,1}$ mode will be cut off. If the guide radius lies between the limits $\lambda/3.4$ and $\lambda/2.6$, the dominant mode only will be propagated. This allows sufficient freedom for practical applications (sidebands plus tuning range may be about 25 per cent of the mid-frequency). Usually the guide is chosen to have a radius close to $\lambda/2.6$ to offer the lowest attenuation. Here again, the dominant mode has the lowest attenuation, with the exception of the anamolous $TE_{0,1}$ mode that produces no currents in the walls of the guide and hence has continually decreasing attenuation as the frequency of operation increases. This is a theoretical advantage only since it has not proved possible to sustain this mode in practice, doubtless because it demands perfect circularity in the guide. At high frequencies, the slightest eccentricity in the guide walls produces instability and the energy transforms to other modes.

The need of circularity in round guides is emphasized by the fact that elliptical guides convert the energy fed to them into two polarizations that have different wavelengths in guide. For this reason it is usually more difficult to match circular guide to terminations than it is to match rectangular guide.

The $TM_{0,1}$ mode in circular guide is widely used, because of its axial symmetry, in rotating joints. As shown in Chap. VII, the transfer from $TE_{1,0}$ mode in rectangular guide to $TM_{0,1}$ mode in circular guide is readily carried out by a probe extending across the rectangular guide and into the circular guide.

80. Excitation and Termination of Waveguides.—Waveguides are usually excited and terminated by coupling the electric or magnetic lines of force of the desired mode by probes or coupling loops, respectively. This form of connection is used for coupling guides to coaxial cables. The electromagnetic radiation within

the guide may also be transferred from one type of waveguide to another by introducing a gradual change in the size or shape of the guide. Finally, the energy within the guide may be radiated by matching the guide to the surrounding space directly or through a flare referred to as a "horn antenna." Figure 132 illustrates typical coupling methods for exciting or terminating a waveguide.

When the termination of a waveguide is considered, the question of standing waves and power transfer arises. It is found that standing waves are produced by terminations and that the standing waves may be compensated by the insertion of reactive

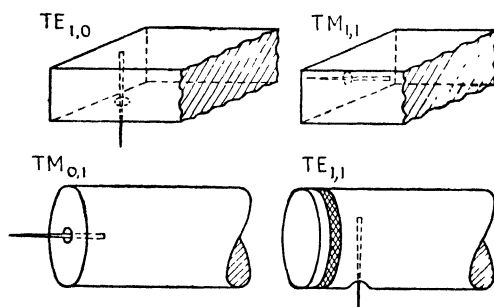



FIG. 132.—Methods of exciting particular modes in rectangular and circular guides.

elements of the proper magnitude and type. A similarity to impedance matching in transmission lines is thus suggested, but the impedance concept in a waveguide is evidently a more complicated matter since it involves field configurations, rather than the simple ratio of voltage to current. Nevertheless, the impedance concept is useful in the design of waveguide systems. In fact, impedance or admittance circle diagrams may be used, and standing-wave ratios and nodal positions may be computed exactly as in transmission lines.

The concept of impedance, as applied to an electromagnetic field, relates to the ratio of the electric field to the magnetic field. It is evident that a very wide variety of impedances may be defined, depending on the components of the electric and magnetic fields considered and on the mode of propagation. Although there is as yet no generally accepted standard among the several proposed definitions, we shall adhere to the so-called "specific wave impedance," which is the ratio of the transverse

electric field to the transverse magnetic field. Evidently the specific wave impedance is different for *TE* and *TM* modes.

The specific wave impedance might be computed from Eqs. (236) through (240), for example, but it proves simpler to express the impedance in terms of the free-space wavelength λ and the wavelength in guide λ_g, since in these terms the impedance has the same value for all modes The specific wave impedance for all *TE* modes is

$$Z_0 = \frac{377\lambda_g}{\lambda} \qquad \text{ohms} \tag{251}$$

The impedance rapidly increases as the cutoff frequency is approached (as the wavelength in guide approaches infinity). Thus the specific wave impedance is not a constant for a particular guide and mode, but varies with the frequency of operation.

In contrast, for *TM* modes the impedance decreases as the frequency approaches cutoff. The specific wave impedance for *TM* modes is

$$Z_0 = \frac{377\lambda}{\lambda_g} \qquad \text{ohms} \tag{252}$$

In the principal or *TEM* mode, the impedance is 377 ohms, since then λ and λ_g are equal. The quantity 377 ohms, is in other words, the specific wave impedance of empty space.

The specific wave impedance of a waveguide may be thought of as a characteristic impedance for the particular guide, mode, and frequency of operation inasmuch as no reflections will occur if the ratio of transverse electric field to transverse magnetic field is maintained at each junction or termination in the guide. But the numerical values expressed by Eqs (251) and (252) are of little value, since the characteristics of probes and lobes and reactive elements in the line cannot be expressed simply in ohms.

Fortunately it is not necessary to deal in specific impedance or admittance values because normalized values will suffice so long as only one mode is present (that is, away from the immediate vicinity of discontinuities). Then the terminating impedance Z_t is expressed in the normalized form $z_t = Z_t/Z_0$, where Z_0 is the characteristic impedance. By confining all computations to normalized values, it makes no difference what

the assumed value of Z_0 may be, and the circle diagrams, plotted in normalized coordinates, then apply. In particular the ratio of the amplitudes of the reflected to direct electric field is given by

FIG. 133.—Smith circle diagram applied to waveguide.

$$\frac{E_2}{E_1} = \frac{k-1}{k+1} \tag{253}$$

and the standing-wave ratio is

$$k = \frac{E_1 + E_2}{E_1 - E_2} \tag{254}$$

From the value of k, using the circle diagram, it is possible to deduce the amount and position of normalized reactance or susceptance required to compensate the standing waves and thus match the termination. Figure 133 shows the circle diagram of a typical case.

81. Reflecting Elements for Impedance Matching in Waveguides.—The reflecting elements used to match impedances in waveguides take several widely used forms: (1) tuning screws, (2) window apertures, inductive or capacitive, (3) series and shunt tees and plungers. Other devices serve the purpose

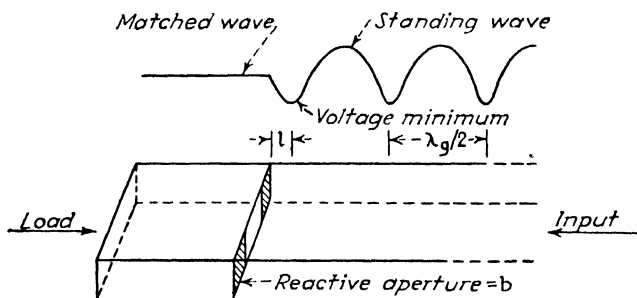



FIG. 134.—Suppression of standing waves by matching aperture in waveguide.

equally well and are used in particular applications. In fact, any obstruction in the guide that does not consume power, and whose longitudinal dimension is small compared with the wavelength in the guide, acts as a reflecting element.

When a reflecting element of normalized susceptance b is introduced into an otherwise uniform waveguide, the standing-

wave ratio produced by the reflecting element is

$$k = \frac{\sqrt{b^2 + 4} + b}{\sqrt{b^2 + 4} - b} \tag{255}$$

The standing-wave ratio is generally measured (in rectangular guide) by a short probe inserted through the top or bottom of the guide parallel to the electric lines of force. The voltage developed on the probe is rectified and indicated on a meter. The ratio of maximum to minimum readings, as the probe is moved through a half wavelength in the guide, is the quantity k^2, assuming a square-law rectifier. From the value of k thus determined, the necessary value of b can be found from Eq. (255). The position of the reflecting element is then found by

$$l = \frac{\pi/2 - \tan^{-1} b/2}{4\pi} \lambda_g \tag{256}$$

where l is the distance from a voltage minimum to the susceptive element and λ_g is the wavelength in guide. Figure 134 illustrates the procedure.

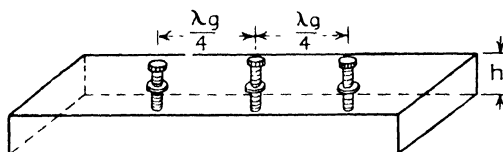



FIG. 135.—Tuning screws.

Tuning Screws.—The tuning screw, as its name suggests, is a threaded cylindrical post, extending through the top of a rectangular guide, parallel to the electric field (Fig. 135). It is used to adjust terminations and junctions to minimum standing-wave ratio. A single post can match a termination only when the spacing between post and termination is critically adjusted. To avoid this spacing requirement, three tuning screws are generally used, spaced one-quarter wavelength (in guide).

The tuning screw introduces a capacitive susceptance when it extends less than a free-space quarter wavelength into the guide. Inductive susceptance is introduced by longer screws, but these are seldom used in practice because of the danger of voltage breakdown and because no advantage obtains over the capacitive screw.

The amount of normalized susceptance introduced is, of course, a function of the operating wavelength, and thus a given setting of the screw is effective over a limited band of frequencies. Screws of large diameter are effective over wider bands than narrow screws. There are no simple expressions for the susceptance introduced by the screw as a function of its dimensions and depth, but fortunately the adjustable attribute of the device makes exact computation unnecessary. Normalized capacitive shunt susceptances up to 5.0 may be obtained by adjusting the length within the guide up to about 65 per cent of the h dimension (height of cross section).

Inductive Apertures.—The inductive aperture (Fig. 136*a*) consists of two thin plates of metal inserted symmetrically

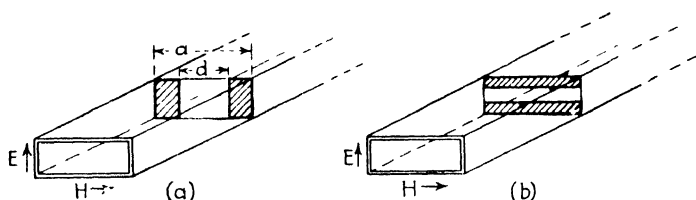



FIG. 136.—Reactive apertures in waveguides: (*a*) inductive and (*b*) capacitive.

across the guide, with their open edges parallel to the electrical field. The inductive nature of the susceptance is evident from the fact that current flows in the aperture walls. The inductance acts in shunt with the characteristic impedance of the guide.

The normalized susceptance is given by

$$b = \frac{-\lambda_g}{a} \cot^2 \frac{\pi d}{2a} \qquad (257)$$

where d and a are the dimensions indicated in Fig. 136*a*, and λ_g is the wavelength in guide. Evidently the smaller the aperture opening d, the greater the susceptance and the higher the standing-wave ratio.

Inductive windows are used when a fixed and relatively large susceptance is required. Tuning screws may be used to make fine adjustments. The inductive window does not lower the breakdown voltage of the guide appreciably and hence is suitable for use in high-power systems.

Capacitive Apertures.—The capacitive aperture (Fig. 136*b*) is similar to the inductive, except that the sides of the opening

are parallel to the magnetic field. The capacitive nature of the susceptance is evident from the fact that the electric lines of force act across the edges of the aperture. For this reason the capacitive window is prone to breakdown and is not used in high-power systems.

The capacitive normalized susceptance is given by

$$b = \frac{4h}{\lambda_g} \log_\epsilon \operatorname{cosec} \frac{\pi d}{2h} \tag{258}$$

where h is the height of the guide and d is the height of the aperture gap. This expression assumes that the aperture plates have infinitesimal thickness. Even moderate thickness of metal has a pronounced effect in a capacitive aperture (in contrast to the inductive aperture), raising the susceptance considerably above the values given by Eq. (258). When the thickness is appreciable, the impedance introduced has a real component, a normalized conductance given by

$$g = \frac{2\pi t}{\lambda_g} d \frac{b_0}{h} \tag{259}$$

where b_0 is the susceptance given by Eq. (258) and t is the thickness of the aperture plates. The normalized susceptance under these conditions is given by

$$b = b_0 + \frac{2\pi t}{\lambda_g}\left(\frac{h}{d} - \frac{d}{h}\right) \tag{260}$$

Other forms of apertures and resonant elements such as rings, disks, posts, centrally located plates, etc., may be used for special purposes.

Tees.—The waveguide tee takes its name from its shape (Fig. 137). It is used (1) as an adjustable impedance matching device in much the same manner as a tuning stub on a transmission line, or (2) as a means of introducing a sharp bend in a waveguide without introducing excessive standing waves. One of the members of the tee is fitted with an adjustable plunger capable of moving one-half wavelength in guide. By adjustment of the plunger the normalized admittance looking into one of the open ends can be made to display capacitive or inductive susceptance as well as conductance, while the other open end displays the characteristic admittance and hence will introduce no reflections when joined to another waveguide of the same cross section. The two principal types shown are (1)

the series tee, in which the side arm is attached to the flat side of the guide and (2) the shunt tee, which is attached to the narrow side.

Twists and Bends.—The rectangular type of waveguide can be twisted and bent (Fig. 138) without serious losses due to reflections, provided that the bend or twist is not too abrupt. If a twist extends over two free-space wavelengths, for example,

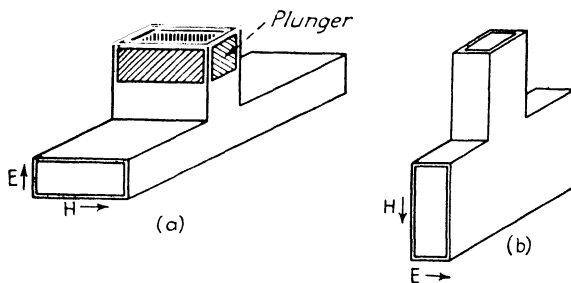



FIG. 137.—Series (*a*) and shunt (*b*) tees

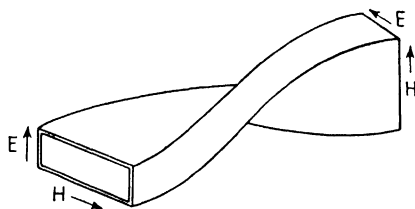



FIG. 138.—Waveguide twist member.

(a) (b)

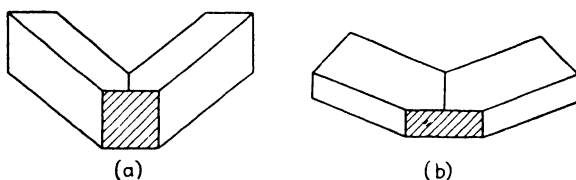



FIG. 139.—Waveguide corners and matching surfaces.

the standing-wave voltage ratio will be of the order of $k = 1.05$. It is necessary, of course, that the cross-sectional dimensions be accurately maintained throughout the twist or bend, which is not always a simple mechanical requirement. Sharp bends (corners) are by nature abrupt; therefore it is necessary, generally, to compensate for reflections by the introduction of additional reflecting elements. Typical corner structures are shown in Fig. 139.

82. Oscillating Circuits and Resonant Cavities.—We have already noted (Sec. 66) that short-circuited or open-circuited segments of transmission line display the properties of tuned circuits. Segments of waveguide, closed at either end, serve the same purpose. Such resonant sections of coaxial line or waveguide are termed "resonant cavities." They are widely used as tuned circuits in radar practice because they are simpler mechanically and more efficient electrically than the lumped-constant tuned circuits used at lower frequencies. In fact the lumped-constant circuit becomes impractical, because of small size and attendant high losses, at frequencies higher than a few hundred megacycles. Resonant cavities are extensively used as oscillators, mixers, wavemeters, and wherever tuning is

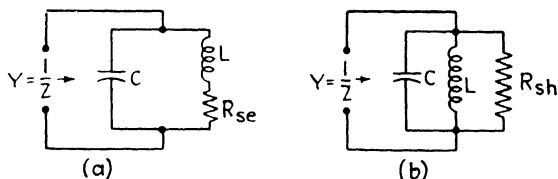



FIG. 140.—Parallel-tuned circuits with series (*a*) and shunt (*b*) resistive elements.

required in r-f equipment. In this section and in the following sections we shall review the properties of tuned circuits and the mechanism of oscillation in resonant cavities.

As a point of departure, it is convenient to state the familiar properties of the dissipative shunt-tuned circuit. The circuit (Fig. 140) consists of an inductance L henries, a capacitance C farads, and a resistance. The resistance is designated as R_{se} ohms if it is in series with the inductance or capacitance (Fig. 140*a*), or R_{sh} ohms if it is in shunt across the circuit (Fig. 140*b*). Resonance occurs when the angular frequency has a particular value ω_r such that the capacitive and inductive reactances are equal, that is, $\omega_r L = 1/\omega_r C$. Hence the resonant angular frequency is

$$\omega_r = \frac{1}{\sqrt{LC}} \qquad \text{cps} \tag{261}$$

The impedance of the circuit is a maximum at resonance and is equal to

$$Z = R_{sh} = \frac{\omega_r^2 L^2}{R_{se}} \qquad \text{ohms} \tag{262}$$

High impedance at resonance is associated with high values of shunt resistance and low values of series resistance.

At frequencies removed from the resonant value, the impedance of the circuit decreases in accordance with the resonance curve. The sharpness of the resonance curve depends on the ratio of the shunt resistance to the reactance of L (or C) at resonance. This ratio, called the "circuit Q," is defined as

$$Q = \frac{R_{sh}}{\omega_r L} = \frac{\omega_r L}{R_{se}} \tag{263}$$

We note that these definitions of Q are consistent with the relationship between R_{se} and R_{sh} expressed in Eq. (262).

The quantity Q is most useful in r-f practice because it can be deduced from selectivity measurements, even if the resistive, inductive, and capacitive elements themselves are not known explicitly. We can trace the relationship of Q to the selectivity characteristic by means of the admittance of the circuit.

$$Y = \frac{1}{R_{sh}} + j\left(\omega C - \frac{1}{\omega L}\right) \tag{264}$$

$$= G + jB \tag{265}$$

where G is the conductance of the shunt resistance, and B is the sum of the susceptances of L and C. The variation of B with frequency is most conveniently expressed in terms of the deviation $\Delta\omega$ from the resonant frequency ω_r. Thus, by reference to Eq. (261) and (264),

$$B = \frac{1}{\omega_r L}\left(\frac{\omega_r + \Delta\omega}{\omega_r} - \frac{\omega_r}{\omega_r + \Delta\omega}\right) \tag{266}$$

The quantity within the parentheses may be written as

$$\frac{2\omega_r\,\Delta\omega + (\Delta\omega)^2}{\omega_r^2 + \omega_r\,\Delta\omega} = \frac{\Delta\omega}{\omega_r}\left(\frac{2\omega_r + \Delta\omega}{\omega_r + \Delta\omega}\right) \tag{267}$$

When the frequency deviation $\Delta\omega$ is small, as it almost always is in practice, this quantity reduces to $2\,\Delta\omega/\omega_r$ and Eq. (266) can be written

$$B = \frac{1}{\omega_r L}\,\frac{2\,\Delta\omega}{\omega_r} \tag{268}$$

Suppose we take a particular frequency deviation

$$\Delta\omega_0 = \frac{\omega_r}{2Q} \tag{269}$$

Then by substituting Eq. (263) and Eq. (269) in Eq. (268) we obtain

$$B = \frac{1}{\omega_r LQ} = \frac{1}{R_{sh}} = G \tag{270}$$

Thus $\Delta\omega_0$ is the frequency deviation at which the conductance and susceptance are equal. The magnitude of the admittance is then $|Y| = 1.414G$, or 41 per cent greater than the resonance value. In impedance terminology, at this frequency deviation

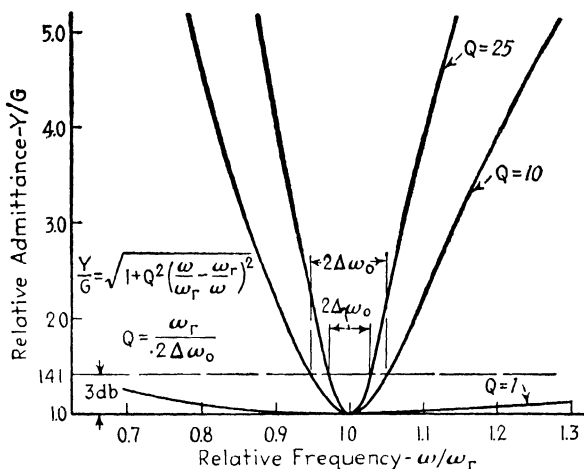



FIG. 141.—Resonance curves of the shunt-tuned circuit, plotted in the admittance form.

the impedance is 70.7 per cent of (or 3 db down from) its resonance value. Thus, rewriting Eq. (269), we find that the circuit Q is equal to the ratio of the resonant frequency ω_r to the total bandwith 2 $\Delta\omega_0$ between the 3-db points on the resonance curve.

$$Q = \frac{\omega_r}{2\,\Delta\omega_0} \tag{271}$$

The relationships between the circuit parameters and corresponding resonance curves are illustrated in Fig. 141.

It is evident that a shunt-tuned circuit, loaded with shunt resistance, may be described in terms of its circuit elements, L, C, and R_{sh}, or equally well in terms of its resonant frequency ω_r, circuit Q, and shunt resistance R_{sh}. For if the resonant frequency is known, the product LC is known. If R_{sh} and Q are

known, L may be determined from Eq. (263), and C is found from the LC product.

In lumped circuits we may approach the circuit either by measuring the inductance, capacitance, and resistance or by measuring the shunt resistance and the relative response at different frequencies, thereby determining ω_r and Q. But in the case of resonant cavities, there is no unique definition of inductance and capacitance; therefore we must perforce measure the resonance curve and the shunt resistance. Fortunately it turns out that the latter measurements are readily performed, whereas inductance and capacitance are difficult to measure, even when uniquely defined, at ultrahigh and superhigh frequencies.

Thus far we have defined Q in terms of inductance (or capacitance) and resistance, but the definition does not apply to resonant cavities for the reasons just stated. Hence we must use another definition of Q that applies to both cases. This definition is

$$Q = \frac{2\pi E_s}{W_d t_p} \tag{272}$$

where E_s is the average value of energy in joules stored in the circuit or cavity, W_d is the power in watts dissipated in losses within the circuit or cavity or to the external load, and t_p is the period of alternation of the applied voltage or current ($t_p = 1/f$ where f is the frequency of operation). We see that this definition agrees with Eq. (263) by the following reasoning. The average energy stored in a tuned circuit is $E_s = \frac{1}{2}LI^2$ where I is the rms value of the current through the inductance L. The average losses in the resistance are $W_d = \frac{1}{2}I^2R_{se}$, and the period t_p is $2\pi/\omega_r$. Substituting these equivalent quantities in Eq. (272) we obtain

$$Q = \frac{2\pi \frac{1}{2}LI^2\omega_r}{\frac{1}{2}I^2R_{se}2\pi} = \frac{\omega_r L}{R_{se}} \tag{273}$$

which is identical with Eq. (263).

Similarly, in terms of energy stored and dissipated, it is possible to find a unique value of Q for each mode of oscillation in any form of resonant cavity (including coaxial forms). The value of shunt resistance cannot be found uniquely, but a conventional form can be adopted and the selectivity and losses of given resonators deduced from it.

83. Energy Storage and Dissipation in Resonant Cavities.—The energy within a waveguide, as we have seen in the development of Eq. (244), depends on the square of the magnetic field intensity **H**. In particular, the energy stored within a resonant cavity is

$$E_s = \frac{1}{8\pi} \int \mathbf{H}^2 \, dv \tag{274}$$

where dv is an element of volume within the cavity and the integration is carried out over the volume of the cavity. Similarly the power dissipated times the period is

$$W_d t_p = \frac{\delta}{8} \int \mathbf{H}^2 |da| \tag{275}$$

where **H** is the tangential field at the cavity wall, da is an element of area on the wall, and δ is the skin depth due to the finite conductivity of the wall. The integration is carried out over the inner surface of the cavity.

Substitution of Eqs. (274) and (275) in Eq. (272) gives the Q of the resonant cavity.

$$Q = \frac{2\int \mathbf{H}^2 \, dv}{\delta \int \mathbf{H}^2 |da|} \tag{276}$$

To compute Q, it is necessary to specify the mode of oscillation, (that is, the distribution of the magnetic intensity H) and to perform the indicated integrations. Fortunately the distributions of H through the volume and over the inner surface are closely related. In fact, the average value of the field at the surface is approximately twice the average value of the field over the volume, and thus we may cancel the H^2 in the numerator by $2H^2$ in the denominator and obtain

$$Q = \frac{\int dv}{\delta \int da} = \frac{v}{\delta a} \tag{277}$$

where v is the volume, a the area of the inner surface, and δ the skin depth, all in the same units. We thus arrive at a value of Q which applies, approximately, for all modes and which is expressed simply in terms of the volume-to-area ratio.

Several aspects of resonator performance are indicated in Eq. (277). (1) We note that large resonators have large Q values, since as the size increases the volume increases faster than the

area, the shape of the resonator remaining unchanged. Physically this is equivalent to the fact that a large resonator can store more energy, for given losses, than a small resonator. (2) We note that very large Q values are possible in resonators, much larger than those found in lumped circuits and exceeding even those found in piezoelectric crystals. In copper at a frequency of 3,000 megacycles, for example, the skin depth δ is about 1.2×10^{-4} cm. A cubic cavity 10 cms on a side ($v/a = 1.6$ cm) thus has a Q value of approximately $(1.6/1.2) \times 10^4 = 13{,}300$. A cubic resonator 50 cms on a side has a Q five times as great, or 65,000.

It is thus evident that, aside from the mechanical convenience of the resonator, the electrical efficiency is high. The selectivity is correspondingly good, which is fortunate in view of the high frequencies employed. A resonator having a Q of 10,000, by Eq. (271), displays at 3,000 megacycles carrier frequency a bandwidth between 3 db points of 300 kc, which is the order of the bandwidth required for a 3-sec pulse. By proper design of the resonator and its load, it is possible to obtain a selectivity just sufficient for the pulse spectrum plus allowance for frequency instability and thus obtain the optimum signal-to-noise ratio.

The shunt resistance of a cavity resonator is also defined in terms of energy dissipation. In conformity to the result obtained for lumped circuits, we define the shunt resistance as

$$R_{sh} = \frac{V^2}{2W_d} \tag{278}$$

where V is the rms voltage developed across the circuit or cavity and W_d is the power dissipated in the circuit or cavity. The energy dissipated is given by Eq. (275) (it should be noted that $1/t_p = f$, the frequency of operation):

$$W_d = \frac{f\delta}{8} \int \mathbf{H}^2|da| \tag{279}$$

It is not possible to define the voltage V uniquely, but in the lower modes it is possible to adopt a simple convention that yields a particular value of V. For example, in a rectangular cavity, with the electric field running from top to bottom of the guide, V is taken as the rms value of voltage between top and

bottom. The voltage thus defined may be related to the magnetic field by the induction law.

$$V = -\frac{1}{c} \int \mathbf{H} \, da_1 \tag{280}$$

where da_1 is an element of area in the plane of the path along which the voltage is taken. Substituting Eqs. (279) and (280) in (278) (and inserting a proportionality constant for emu units), we obtain

$$R_{sh} = \left[16\pi^2 \frac{(\int \mathbf{H} \, da_1)^2}{\lambda^2 \int \mathbf{H}^2 |da|} \right] \frac{\lambda c}{\delta} \qquad \text{emu} \tag{281}$$

The quantity in brackets is a form factor dependent only on the shape of the cavity and the mode of oscillation. The quantity outside the brackets has, at 10-cm a wavelength ($\lambda = 10$ cm, $\delta = 1.2 \times 10^{-4}$ cm (copper), and $c = 3 \times 10^{10}$ cm per sec), a value of

$$\frac{\lambda c}{\delta} = 2.5 \times 10^{15} \text{ abohms} = 2.5 \qquad \text{megohms} \tag{282}$$

Thus, when the form factor is approximately unity, as it is in practical cases, the shunt resistance of the cavity is measured in megohms. We note that the shunt resistance varies not with the size of the resonator, but with its shape only.

From the approximate expressions in Eq. (277) and (282) it is thus possible to obtain qualitative information on two of the three essential descriptive quantities, Q and R_{sh}. For information on the third quantity, the resonant frequency ω_r, we may resort to measurements. But then we find a multiplicity of resonant frequencies corresponding to the various modes of oscillation. To deal with this situation, we must compute the field configurations. This computation will not only isolate particular modes of interest but will permit more accurate computation of Q and R_{sh}.

84. Modes in Cavity Resonators. Rectangular Form.—Consider a resonator in the form of a box with rectangular sides (Fig. 142). The dimensions in the x, y, and z coordinates are, respectively, a, b, and c cms. The cavity is excited with an transverse electric vector in the xy plane, as in the TE mode in waveguide (Fig. 125), and a wave tends to be propagated along

the Z axis. The side wall blocks and reflects this wave returning it along the Z axis to the opposite wall, where it is reflected again. In this manner a standing wave is set up along the Z axis, which, in conjunction with the standing waves along the X and Y axis, constitutes the electric field distribution within the resonator.

The field distribution is found in a manner exactly analogous to the derivation of the TE modes in rectangular guide, taking into account the effect of the closed space along the Z axis. We recall that the vector wave equation, Eq. (222), is satisfied

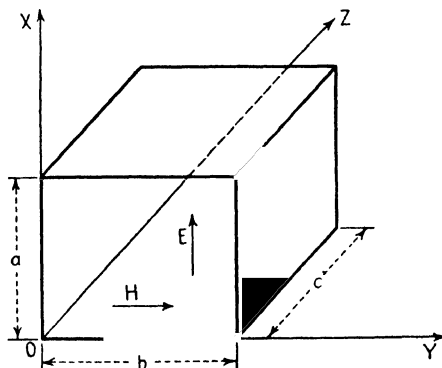



FIG. 142.—Basic geometry of cubical resonator.

by sinusoidal variations in space along each axis. Moreover, the number of half-wave variations along each axis is not limited but may have values l, m, and n along the X, Y, and Z axes, respectively. The boundary conditions must be satisfied at all the walls, that is, vanishing tangential electric and normal magnetic fields (assuming perfect conductivity). To assure a normal electric field, the sinusoidal variation must be of the form $\sin \pi lx/a$, along the X axis, $\sin \pi my/b$, along the Y axis, and $\sin \pi nz/c$, along the Z axis, where the integers l, m, and n may have any value, 0, 1, 2, 3, etc.

At resonance the quantity l/a indicates the number of sinusoidal variations per unit length along the x axis, m/b those along the Y axis, and n/c those along the Z axis. Moreover, at resonance the vector sum of these quantities equals the inverse of the resultant half wavelength. Thus

$$\frac{2}{\lambda_0} = \sqrt{\left(\frac{l}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{n}{c}\right)^2} \qquad (283)$$

from which the resonant wavelength becomes

$$\lambda_0 = \frac{2}{\sqrt{(l/a)^2 + (m/b)^2 + (n/c)^2}} \tag{284}$$

This is the free-space wavelength corresponding to a particular mode, designated as the $TE_{l,m,n}$ mode.

We note that the electric field must vanish at the Z axis end plates; and thus a uniform field along the Z axis is ruled out, and the value $n = 0$ is thus not permitted. Moreover, either l or m may be zero, but not both simultaneously. Except for these restrictions, l, m, and n may have any values and each combination leads to a different value of resonant wavelength, in accordance with Eq. (284). It is evident that if the wavelength that excites the resonator is small compared with the dimensions, corresponding high values of l, m, and n exist and, in fact, several modes of oscillation may be excited simultaneously. To minimize this effect, it is customary to choose the dimensions so that the resonator is almost, but not quite, a perfect cube.

A particular form of rectangular resonator is that in which $b = c$, the a dimension is not specified, and which is operated in the $TE_{1,1,0}$ mode. Then, by Eq. (284),

$$\lambda_0 = 1.41b \qquad \text{cms} \tag{285}$$

The Q and R_{sh} values for this case may also be stated in fairly simple form.

$$Q = \frac{b}{\delta}\frac{1}{1 + b/2a} \tag{286}$$

$$R_{sh} = 42.5\frac{a}{\delta}\frac{1}{1 + b/2a} \tag{287}$$

where δ is the skin depth in centimeters, and a, b, and c are the dimensions in centimeters of Fig. 142.

Cylindrical Resonators.—The computation of the modes in cylindrical cavity resonators (Fig. 143) is based on the modes of circular waveguides. The subscripts m and n refer, respectively, to the number of radial and perimetric variations of the field across the circular cross section. An additional subscript l denotes the number of field variations along the Z axis, within the height h of the resonator. As in the case of rectangular resonators, the quantity l/h indicates the number of sinusoidal

variations, at resonance, per unit length along the Z axis. The vector sum of this quantity and the inverse half wavelength at cutoff of the corresponding circular guide equal the inverse resonant half wavelength of λ_0 within the cavity. Thus, referring to Eq. (241), we can write

$$\frac{2}{\lambda_0} = \sqrt{\left(\frac{l}{h}\right)^2 + \left(\frac{x_{m,n}}{\pi a}\right)^2} \tag{288}$$

Resonant wavelength may be written explicitly by rearranging Eq. (287)

$$\lambda_0 = \frac{2}{\sqrt{(l/h)^2 + (x_{m,n}/\pi a)^2}} \tag{289}$$

where a is radius of cavity, h is height, and $x_{m,n}$ is the mth root of the Bessel function (TM modes) or its derivative (TE modes) of order n. The values of $x_{m,n}$ are given in Table VI, page 190. Since the tangential field must vanish at $z = 0$ and $z = h$, it follows that there must be at least one sinusoidal variation along z; and thus $l = 0$ is not permitted, but any nonzero value of l is allowable. As in circular guides, n may have any integer value, including zero, and m any integer value except zero.

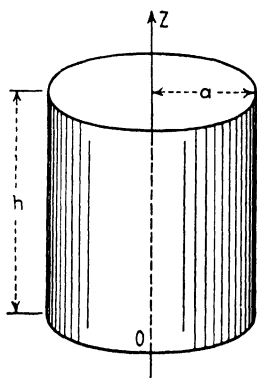



FIG. 143.—Basic geometry of cylindrical resonator.

The Z and R_{sh} values of cylindrical resonators are complicated functions of $x_{m,n}$, h, a, and δ (the skin depth) when expressed in general terms for any mode. One particular mode of interest is the $TM_{0,1,0}$, in which the electric field is wholly parallel to the z axis. The cavity Q is then

$$Q = \frac{a}{\delta} \frac{1}{1 + 2a/h} \tag{290}$$

and the shunt resistance is

$$R_{sh} = 72 \frac{\lambda_0 h}{\delta a} \frac{1}{1 + a/h} \tag{291}$$

The resonant wavelength of this mode, by reference to Eq. (287) and Table VI, is found to be

$$\lambda_0 = \frac{2\pi a}{2.4} = 2.6a \tag{292}$$

Spherical Resonators.—The sperical resonator (Fig. 144) has no analogue in waveguides. The modes are found by expressing

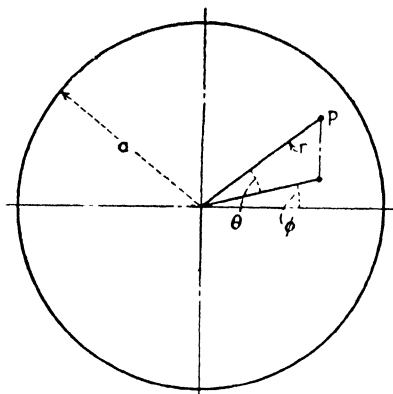



FIG. 144.—Basic geometry of spherical resonator.

Maxwell's equations in spherical coordinates (r, ϕ, and θ) and by solving the corresponding spatial-wave equation. The solutions are found to be products of half-order Bessel functions ($m + \frac{1}{2}$ order) of argument r, associated Legendre functions of mth degree and nth order of argument $\cos \theta$, and sinusoidal functions of argument $n\phi$. Fortunately this formidable array of mathematical quantities reduces to sine and cosine terms in particular low-order modes. It will be noted that only two subscripts, m and n, are required.

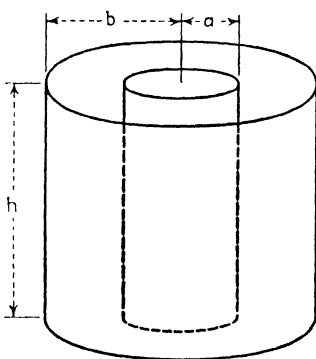



FIG. 145.—Geometry of coaxial resonator.

The essential quantities describing the dominant mode in a spherical resonator of radius a are

$$\lambda_0 = 2.28a \tag{293}$$

$$Q = \frac{0.318\lambda_0}{\delta} \tag{294}$$

and

$$R_{sh} = \frac{104.4\lambda_0}{\delta} \tag{295}$$

Coaxial Resonators.—A type of resonator much used in practical applications is the coaxial form (Fig. 145), which is resonant when the height of the resonator is a half wavelength.

$$\lambda_0 = 2h \tag{296}$$

The electric field, in this dominant mode, is wholly radial and the magnetic field is wholly concentric. The general expression for the cavity Q is

$$Q = \frac{\lambda_0}{\delta} \frac{1}{4 + \dfrac{h(1 + b/a)}{b \log_\epsilon b/a}} \tag{297}$$

The highest Q is obtained, other quantities being equal, when

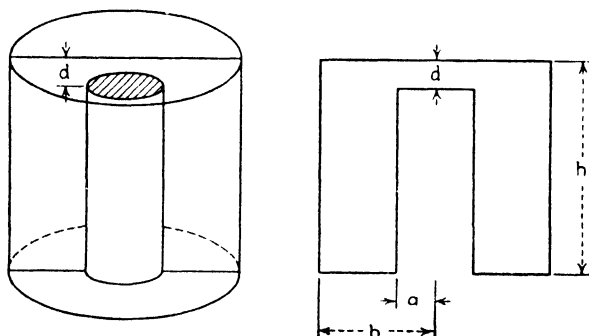



FIG. 146.—Reentrant coaxial resonator.

the ratio of outer radius to inner radius has the values $b/a = 3.6$. The shunt resistance is highest, however, when $b/a = 9.2$.

A modified form of the coaxial resonator (Fig. 146) has its center post connected only at one end. This shape permits variation of the electric field from one side of the cavity around to the other, so that the half-wave resonant length is in effect folded back on itself. This permits resonance to occur when the cavity is one-quarter wavelength long. The lumped capacitance between the open end of the center post and the opposite face may be adjusted to tune the cavity to resonance. An approximate expression for the resonant wavelength is

$$\lambda_0 = 2\pi \cdot \sqrt{\frac{2ha^2}{d}\left(\log_\epsilon \frac{b}{a}\right)} \tag{298}$$

85. Excitation of Cavity Resonators.—Cavity resonators are excited by means similar to those employed with waveguides, that is, by probes coupled to the electric lines of force, or loops coupled to the magnetic lines. Another form of excitation, used with the quarter-wave coaxial resonator (Fig. 147), is electronic in nature. The center conductor is hollow and covered at its inner end by a mesh that faces a similar mesh in the opposite face. A stream of velocity-modulated electrons is passed through the meshes and induces between them a varying electric potential. If the induced frequency is resonant, the electric and magnetic fields within the resonator build up in amplitude until the losses in the walls and those abstracted from the cavity just equal the power supplied by the electron beam. This mode of operation is the basis of the klystron oscillator and amplifier, described in Chap. VII.

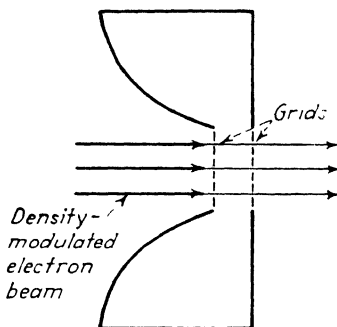



FIG. 147—Reentrant resonator with grid surfaces excited by passage of electron beam

Bibliography

GUILLEMIN, E. A.: "Communication Networks," vols. I and II, John Wiley & Sons, Inc., New York, 1931 and 1935.

BRAINERD and Others: "U-h-f Techniques," D. Van Nostrand Company, Inc., New York, 1942.

REICH, H. J.: "Theory and Applications of Electron Tubes," McGraw-Hill Book Company, Inc., New York, 1939.

SLATER, J. C.: "Microwave Transmission," McGraw-Hill Book Company, Inc., New York, 1942.

TERMAN, F. E.: "Radio Engineers' Handbook," McGraw-Hill Book Company, Inc., New York, 1943. See particularly Sec. 3, pp. 135–251.

SKILLING, H. H.: "Transient Electric Currents," McGraw-Hill Book Company, Inc., New York, 1937.

SARBACHER and EDSON: "Hyper and Ultra-high Frequency Engineering," John Wiley & Sons, Inc., New York, 1943.

RAMO and WHINNERY: "Fields and Waves in Modern Radio," John Wiley & Sons, Inc., New York, 1944.

BARROW, W. L.: Transmission of Electromagnetic Waves in Hollow Tubes of Metal, *Proc. I.R.E.*, **24**, 1298 (October, 1936).

SOUTHWORTH, G. C.: Some Fundamental Experiments with Waveguides, *Proc. I.R.E.*, **25**, 807 (July, 1937).

SOUTHWORTH, G. C.: Hyper Frequency Wave Guides, *Bell Sys. Tech. Jour.*, **15**, 284 (April, 1936).

CARSON, MEAD, and SCHELKUNOFF: Hyper Frequency Waveguides—Mathematical Theory, *Bell Sys. Tech. Jour.*, **15**, 310 (April, 1936).

HANSEN, W. W.: A Type of Electrical Resonator, *J. Applied Phys.*, **9**, 654 (October, 1938).

HANSEN, W. W.: On the Resonant Frequency of Closed Concentric Lines, *J. Applied Phys.*, **10**, 38 (January, 1939).

CONDON, E. U.: Forced Oscillations in Cavity Resonators, *J. Applied Phys.*, **12**, 129 (February, 1941).

BARROW and MIEHER: Natural Oscillations of Electrical Cavity Resonators, *Proc. I.R.E.*, **28**, 184 (April, 1940).

CHAPTER IV

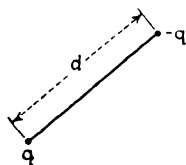
RADIO FREQUENCY FUNDAMENTALS (CONTINUED) RADIATORS AND REFLECTORS, PROPAGATION, AND TARGETS

In this chapter we treat the external aspects of the r-f (radio-frequency) portions of a radar system, that is, those relating to the emission of the signal, propagation through space to the target, absorption and reradiation at the target, propagation back to the radar, and absorption by the radar receiving system.

86. Forms of Radiating Structures.—The radiating structures considered in radar take five basic forms: the isotropic radiator, the infinitesimal dipole, the linear radiator, the horn radiator, and the loop radiator. The first two are of theoretical interest only; the others are practical radiators. In this section we consider the radiating properties of these forms. The formation of beams by arrays or reflectors is treated in Secs. 90 and 91.

The isotropic or "point source" radiator is a theoretical structure that radiates an equal density of energy in all directions. It serves as a convenient reference for considering the directive properties of practical radiators and reflectors. The isotropic radiator, as we shall see in the following discussion, cannot be realized in practice except by assembling a very large number of linear or horn radiators (cf. a point source of light, which constitutes a very large number of radiating atoms, with random phases and orientations).

FIG. 148.—Electric dipole.

The infinitesimal dipole radiator is likewise a theoretical structure of interest because its properties are simple to compute and closely approximate those of real linear radiators. The dipole is defined as a combination of two electric changes, equal in size but opposite in sign, located very close together (Fig. 148). A separating force must be assumed to keep the two charges apart. A static dipole (unchanging charges) does not radiate. But if the magnitude and sign of the two charges change periodically, radiation occurs. The radiation from such an oscillating

dipole is zero along the line connecting the two charges and a maximum at right angles to this line. The distribution is very similar to that of the half-wave linear radiator which is, in fact, often loosely called a "half-wave dipole."

87. Properties of the Infinitesimal Oscillating Dipole.—The oscillating dipole may be represented by a very short wire carrying alternating current. The dipole is oriented in spherical coordinates along the x axis, as shown in Fig. 149. The current

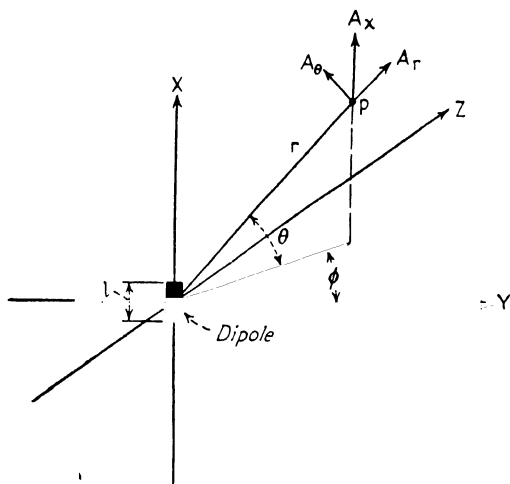



FIG. 149.—Coordinates for computing electromagnetic field surrounding a dipole.

density in the wire is denoted by j_x, where the subscript indicates that the direction of the current is along the x axis (along the wire). The charge density is denoted by ρ. This density represents the alternating accumulation of charge at the ends of the wire under the effects of the alternating current.

To examine the radiating properties of the dipole, we must compute the electric and magnetic fields that surround the dipole as a result of the sinusoidal variation (with time) of the charge and current. As is customary in the solution of electromagnetic field problems in free space, the fields are determined from potentials, that is, quantities[1] which yield the required fields when they are differentiated.

[1] The most familiar potential of this sort is the electrostatic potential $\phi = q/r$, which is the potential caused by a static point charge q at a distance

There are two potentials required, a scalar potential V and a vector potential **A**, such that the magnetic and electric fields are obtained by vector differentiation.

$$\mathbf{H} = \nabla \times \mathbf{A} \tag{299}$$

and

$$\mathbf{E} = -\nabla V - \frac{1}{c}\frac{\delta \mathbf{A}}{\delta t} \tag{300}$$

As a consequence of the above and of Maxwell's equations, the electrodynamic potentials V and **A** are found to be

$$V = \int \frac{\rho(t - r/c)}{r}\,dv \tag{301}$$

$$\mathbf{A} = \frac{1}{c}\int \frac{j(t - r/c)}{r}\,dv \tag{302}$$

Equation (301) states that the scalar potential V is to be found by integrating the charge density ρ over the volume v in which it is found and that the potential decreases with the first power of the distance r from the antenna. The factor $(t - r/c)$ indicates that the variation of the potential with time is propagated through space at a velocity c. This is equivalent to the statement that the value of potential at a particular time and place is the same as that r nearer the source r/c sec earlier in time.

Similarly the vector potential **A** is found by integrating the current density j through the volume v in which it is found. The potential varies inversely with the distance r, and the potential at a particular place and time is the same as that at a place r nearer the antenna and occurring r/c earlier in time. Thus, the change in the potential **A**, like V, is propagated outward from the source at a velocity c.

In the case of the oscillating dipole (Fig. 149) we determine, first, the component of the potential **A** along the antenna axis, by integrating the current density j_x,

$$j_x = \frac{i}{dy\,dz} \tag{303}$$

---

r. The electric field at this distance is $E_r = d\phi/dr = -q/r^2$, which we recognize as a statement of Coulomb's law that the electric field (the force on a unit charge) E_r, varies inversely as the square of the distance from the charge causing the field.

where i is the current in the wire, and $dy\,dz$ is the cross section of the volume element. The current is assumed to be sinusoidal with time at angular frequency ω.

$$i = I \sin \omega t \tag{304}$$

and the element of volume dv is

$$dv = dx\,dy\,dz \tag{305}$$

Substituting these quantities in Eq. (302), we obtain for A_x, the x component of the vector potential **A**

$$A_x = \frac{1}{c} \int_{-l/2}^{l/2} \frac{I \sin \omega(t - r/c)\,dx}{r} \tag{306}$$

$$= \frac{Il}{cr} \sin \omega \left(t - \frac{r}{c}\right) \tag{307}$$

where l is the length of the dipole. We must now find the components of A_x in spherical coordinates, as in Fig. 149. Since x axis is perpendicular to the ϕ plane, $A_\phi = 0$. At the point p, A_x divides into components A_r and A_θ in accordance with the sine and cosine of θ. Thus

$$A_r = A_x \cos \theta \tag{308}$$

and

$$A_\theta = A_x \sin \theta \tag{309}$$

We are now in a position to find the magnetic field by taking the curl of **A** in accordance with Eq. (299). The curl of a vector in spherical coordinates is a highly complex expression, but the result is simple. As we would expect from the fact that the current is along the x axis, the magnetic field is confined to concentric circles about the x axis, that is, along the ϕ coordinate, and the other magnetic components are zero. Substitution of Eqs. (308) and (309) in Eq. (299) shows this to be true.

$$H_r = 0 \tag{310}$$

$$H_\theta = 0 \tag{311}$$

$$H_\phi = \frac{Il}{cr} \sin \theta \left[\frac{\omega}{c} \cos \omega \left(t - \frac{r}{c}\right) + \frac{1}{r} \sin \omega \left(t - \frac{r}{c}\right)\right] \tag{312}$$

The electric field may be found from the spherical components of **A** by Eq. (300), which may be written in the form

$$\mathbf{E} = -\frac{1}{c}\frac{\delta \mathbf{A}}{\delta t} + c\nabla \cdot \int \mathbf{A}\,dt \tag{313}$$

The substitution of Eqs. (308) and (309) in (313) produces

$$E_\phi = 0 \tag{314}$$

$$E_r = \frac{2Il}{\omega r}\cos\theta\left[\frac{\omega}{cr}\sin\omega\left(t-\frac{r}{c}\right) - \frac{1}{r^2}\cos\omega\left(t-\frac{r}{c}\right)\right] \tag{315}$$

$$E_\theta = \frac{Il}{\omega r}\sin\theta\left[\frac{\omega^2}{c^2}\cos\omega\left(t-\frac{r}{c}\right) + \frac{\omega}{cr}\sin\omega\left(t-\frac{r}{c}\right) - \frac{1}{r^2}\cos\omega\left(t-\frac{r}{c}\right)\right] \tag{316}$$

We note that there is a radial component of electric field as well as a transverse component along the θ coordinate.

Having thus found the magnetic and electric fields in the space surrounding the dipole, we must distinguish between two regions, one near the radiator (within a wavelength), the other removed (many wavelengths distant). As is well appreciated the induction field predominates near the antenna, where energy is alternately emitted and substantially reabsorbed by the antenna. At greater distances, however, this induction field diminishes to the vanishing point and the radiation field remains. This field is not reabsorbed by the antenna, but it is permanently detached, traveling through space as an electromagnetic wave.

The induction and radiation fields may be identified in the fields around the dipole by reference to the terms within the brackets in Eqs. (312), (315), and (316). The terms containing r and r^2 in the denominator, which constitute the induction field, become vanishingly small at great distances (r greater than several wavelengths). This is the situation generally met in radar detection, even with near-by targets, because of the short wavelength employed. Thus, in the limit where these terms vanish, there are only two components of field.

$$E_\theta = \frac{Il}{cr}\sin\theta\left[\frac{\omega}{c}\cos\omega\left(t-\frac{r}{c}\right)\right] \tag{317}$$

and

$$H_\phi = \frac{Il}{cr}\sin\theta\left[\frac{\omega}{c}\cos\omega\left(t-\frac{r}{c}\right)\right] \tag{318}$$

These are the components of the radiation field. We note that the electric and magnetic fields have the same magnitude, but occur along the mutual perpendicular axes of θ and ϕ.

We note further that the magnitude of the fields is proportional to the current amplitude I and to the length l of the antenna (remembering, however, that this is small compared to a wavelength and to the distance of observation r). We note also that the magnitude of the fields increases with the angular frequency, and that the fields are propagated away from the source at the velocity c. Finally we note that the electric and magnetic fields are zero along the axis of the radiator, since $\sin \theta = \sin 0 = 0$, and that the field is a maximum at right angles to the antenna,

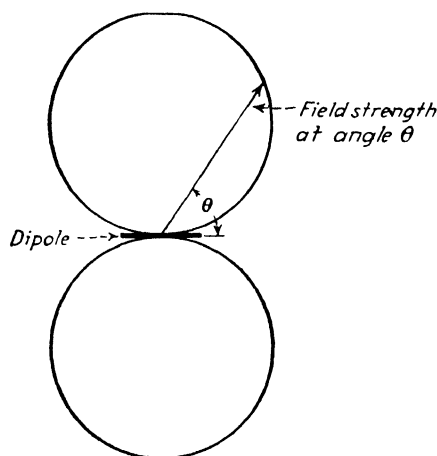



FIG. 150.—Polar diagram of electric-field strength surrounding an elemental dipole.

since $\sin \theta$ then equals $\sin 90° = 1$. The distribution of the field at intermediate angles is that of a sinusoid. Usually the distribution is plotted in polar coordinates. As shown in Fig. 150, the polar diagram of the electric field strength in the plane of the dipole is a figure eight. These properties of a dipole are generally well known to technicians, but their foundation in theory is considered important enough to warrant the foregoing brief review.

Power Radiated from a Dipole.—The radiation fields of the dipole, given by Eqs. (317) and (318), give rise to a radial flow of energy, in accordance with Poynting's vector

$$\mathbf{S} = \frac{c}{4\pi} (\mathbf{E} \times \mathbf{H}) \tag{319}$$

The vector $\mathbf{S}$ represents the instantaneous power flow per unit

area, (for example, watts per square meter), and its direction is at right angles to **E** and **H**. Since the radiation fields of the dipole E_θ and H_ϕ are at right angles to each other, the magnitude of **S** is simply their product (times $c/4\pi$) and its direction is at right angles to θ and ϕ, that is, radially outward. Thus, from Eqs. (317) and (318)

$$|\mathbf{S}| = \frac{I^2l^2}{4\pi cr^2} \sin^2 \theta \frac{\omega^2}{c^2} \cos^2 \omega \left(t - \frac{r}{c}\right) \tag{320}$$

This is the instantaneous power density along the direction θ. The average power density radiated is evidently maximum

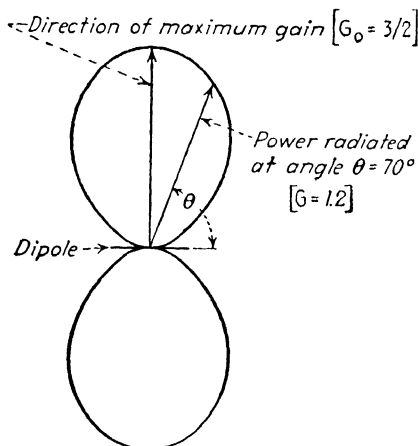



FIG. 151.—Polar diagram of power radiated from an elemental dipole, formed by "squaring" Fig. 150.

along $\theta = 90$ deg, that is, at right angles to the dipole axis. Figure 151 shows the power density distribution in polar coordinates.

In addition to knowing the power density along a particular direction θ, we are also interested in the total power radiated, which is found by integrating Eq. (320) over the surface of a sphere concentric on the source. The element of area on a sphere is

$$da = 2\pi r^2 \sin \theta \, d\theta \tag{321}$$

$|\mathbf{S}|$ in Eq. (320) is multiplied by this element and integrated from $\theta = 0$ to $\theta = \pi$ to find the total power radiated.

$$P = \int \mathbf{S} \cdot da = \frac{I^2 l^2 \omega^2}{2c^3} \cos^2 \omega \left(t - \frac{r}{c}\right) \int_0^\pi \sin^3 \theta \, d\theta \quad (322)$$

$$= \frac{2}{3} \frac{I^2 l^2 \omega^2}{c^3} \cos^2 \omega \left(t - \frac{r}{c}\right) \quad (323)$$

This is the total power radiated at a particular instant of time t. Usually we are interested in the average power radiated over a full period of the wave. To find the average power, we take the average value of $\cos^2 \omega(t - r/c)$, which is equal to ½, and substitute it in Eq. (323).

$$P_{av} = \frac{I^2 l^2 \omega^2}{3c^3} \quad (324)$$

This is the average power radiated by a dipole of length l at frequency ω when it carries a current of maximum amplitude I amp.

Gain of a Dipole.—The dipole, it will be noted from Fig. 151, is a directive radiator, albeit slightly so. Hence it is of interest to define and compute the gain of the dipole relative to an isotropic radiator. The gain of a radiator along a particular direction θ is defined as the ratio of the Poynting vector along that direction to the average value of the Poynting vector over the surface of a concentric sphere. The average Poynting vector is precisely that which would be generated in all directions by an isotropic source of the same power.

The maximum gain of a dipole (Fig. 151) occurs along the direction $\theta = 90$ deg. The average value of the Poynting vector is found by averaging the polar radii of the power distribution in Fig. 151. The average radius is found to be two-thirds the maximum radius. Hence the maximum gain is

$$G_0 = \frac{S_{max}}{S_{av}} = \frac{1}{2/3} = 1.5 \quad (325)$$

Thus the maximum gain of the infinitesimal dipole is 50 per cent greater than the isotropic radiator. The gain in other directions is given in Fig. 151.

Dipole Radiation Resistance.—The radiation resistance of a radiator is a fictitious resistance representing the loss of power in radiation. It is defined as the quotient of the average power radiated by the rms current squared at the point where the antenna

is fed, which in the case of the dipole is at the center. Thus

$$R_{rad} = \frac{P_{av}}{I_{rms}^2} \tag{326}$$

To compute the radiation resistance of a dipole we refer to Eq. (324) for the average power, noting that the current I is the maximum amplitude equal to $1.41 I_{rms}$. Substituting $I^2 = 2I_{rms}^2$ and Eq. (324) in Eq. (326), we obtain for the radiation resistance of a dipole

$$R_{rad} = \frac{2}{3}\frac{l^2\omega^2}{c^3} \quad \text{statohms} \tag{327}$$

$$= \frac{20\omega^2 l^2}{c^2} \quad \text{ohms} \tag{328}$$

88. Properties of the Half-wave Linear Radiator.—The half-wave linear radiator, widely used in radar as well as in other shf and uhf systems, is closely related to the infinitesimal dipole discussed in the preceding section. The important difference is distribution of the current along the dipole axis. In the infinitesimal dipole the current is considered to be uniform, at any instant of time, along the length of the radiator. In the finite half-wave radiator the current must necessarily be zero at the ends of the wire and have its maximum value at the center. The distribution of the current along the radiator may be assumed to be sinusoidal without introducing appreciable error in the deduced properties. Therefore, the current in the half-wave radiator is taken as

$$i = I_0 \cos \frac{2\pi x}{\lambda} \sin \omega t \tag{329}$$

where I_0 is the maximum current at the center of the radiator and x is distance along the radiator measured from its center. When the magnetic potential is computed for this current and the magnetic field found by vector differentiation, the radiation component is found to be

$$H_\phi = \frac{2I_0}{cr}\frac{\cos(\pi/2 \cos\theta)}{\sin\theta} \cos\omega\left(t - \frac{r}{c}\right) \tag{330}$$

and the electric field is

$$E_\theta = \frac{2I_0}{cr}\frac{\cos(\pi/2 \cos\theta)}{\sin\theta} \cos\omega\left(t - \frac{r}{c}\right) \tag{331}$$

Comparison of these fields with those for the infinitesimal dipole, Eqs. (317) and (318), shows them to be identical except for the factor that represents the polar distribution of the field. *The dipole distribution is given by* $\sin \theta$; *the* **half-wave radiator** distribution by $\cos (\pi/2 \cos \theta)/\sin \theta$. In the *equatorial plane* ($\theta = 90$ deg), both factors are the same, that is, unity, which

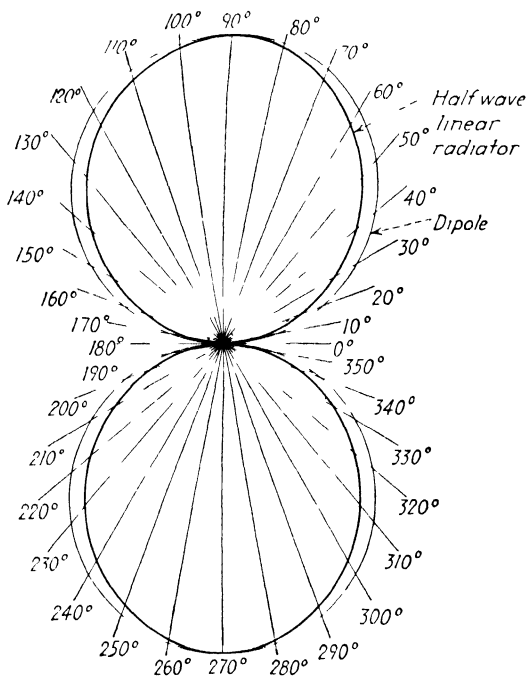



FIG 152—Polar diagram of the electric field of a real half-wave linear radiator, compared with that of an elemental dipole.

indicates that the field in the direction of maximum gain is the same in the dipole and half-wave radiator when both carry the same current. At other angles the two expressions are very nearly the same, as shown in Fig. 152.

Power Radiated by a Half-wave Radiator.—Formation of the Poynting vector from H_ϕ and E_θ, Eqs. (330) and (331), gives the instantaneous power density radiated at the angle θ.

$$|S| = \frac{4I_0^2}{c^2r^2}\left[\frac{\cos (\pi/2 \cos \theta)}{\sin \theta}\right]^2 \cos^2 \omega\left(t - \frac{r}{c}\right) \qquad (332)$$

Figure 153 shows the polar diagram of the radiated power, averaged over a full cycle of the wave.

From Fig. 153 we can compute the maximum gain by finding the average radius of the polar figure and by comparing it with the maximum radius. The result is $S_{av} = 0.61 S_{max}$. The maximum gain of the half-wave radiator is then

$$G_0 = \frac{S_{max}}{S_{av}} = \frac{1}{0.61} = 1.64 \qquad (333)$$

which is slightly higher than that of the infinitesimal dipole.

FIG. 153.—Power polar diagrams of half-wave radiator and dipole.

The radiation resistance of the half-wave radiator, defined by Eq. (326), may also be expressed in terms of the specific impedance of space, 377 ohms, and of the average value of the polar distribution function.

$$R_{rad} = \frac{377}{\pi} \overline{\left(\frac{\cos \pi/2 \cos \theta}{\sin \theta} \right)^2} \qquad (334)$$

where the bar indicates the average value of the quantity over all values of θ. We have already found from Fig. 153 that this average value is 0.61; hence

$$R_{rad} = \frac{377}{\pi} 0.61 = 73.5 \qquad \text{ohms} \qquad (335)$$

which is the familiar figure for the half-wave radiator.

FIG. 154. Polar electric-field diagrams of long linear radiators.

Long Linear Radiators.—Resonant linear radiators longer than a half wavelength are seldom used in directive systems because the polar distribution of radiated power is not suitable. Figure 154 shows several typical examples. Generally there are, in the diagram, a number of loops equal to twice the number

of half wavelengths in the radiator, and usually there are two or more equally prominent loops. To produce only one major loop and to minimize the rest, it is usual to use an array of half-wave radiators, which are fed in phase, as described in Sec. 90. The same result can be obtained from a continuous reflector illuminated from a half-wave source.

89. Waveguide Radiators.—When a waveguide is open at one end, the energy propagated to the opening is partly reflected, due to the impedance mismatch, and partly radiated (Fig. 155). This is the basic action of the waveguide radiator. The polar distribution of the radiated energy and of the gain of the radiator depends on the shape of the opening and on the mode of propagation within the guide. A single mode of low order is preferred to confine the radiated energy to a single radiation pattern.

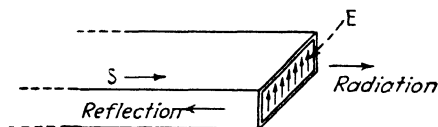



FIG. 155.—Waveguide radiator.

Generally transverse electric modes are used, $TE_{1,0}$ in rectangular guides and $TE_{1,1}$ in circular guides.

The electric and magnetic fields at the open end of a guide carrying *TM* energy are very similar to those within the induction region of a dipole radiator. There are transverse electric and magnetic components and a longitudinal electric component, corresponding with the fields of Eqs. (316), (312), and (315), respectively. Several wavelengths beyond the open end of the guide, the field loses its longitudinal component and becomes a purely transverse (*TEM*) wave. Similarly, the *TE* mode at the open end of a waveguide resembles the fields close to a magnetic dipole (loop radiator). In this case the longitudinal component is magnetic and it disappears within a few wavelengths of the radiator, leaving a *TEM* wave indistinguishable from that radiated by the *TM* mode. The preference of *TE* modes in practice is dictated by somewhat superior directive characteristics, particularly by freedom from back radiation caused by currents along the outer surface of the guide wall.

The gain of a waveguide radiator is proportional to the ratio of the area of the opening to the square of the free-space wave-

length. But, as we recall from Sec. 79, the cross-sectional dimensions of a waveguide must not exceed one-half the free-space wavelength, approximately, if higher order modes are to be avoided. The rectangular guide for a wavelength λ generally has dimensions not greater than λ wide by $\lambda/2$ high, or a cross-section area of $\lambda^2/2$. The gain of such a radiator is somewhat greater than that of a dipole, and the polar distribution, in the electric plane, is similar to the figure-eight pattern of a dipole. The advantage of the waveguide radiator over the dipole is its greater tuning range, the range being limited only by the usual limitations in the waveguide dimensions to avoid propagation by higher order modes. Moreover, this form of the radiator is inherently adapted to waveguides, and hence it is indicated when waveguides are used to conduct the energy from the transmitter to the radiator.

The gain of the waveguide radiator may be computed from consideration of the transverse electric field across the open face of the guide. This alternating electric field may be considered as that corresponding to a large number of alternating charges, or dipoles, each having a directive pattern of cardioid shape. The summation of these patterns produces the over-all radiation pattern. The gain of such an array of dipoles, relative to an isotropic radiator, is

$$G_0 = \frac{4\pi}{\lambda^2} \frac{|\int f\,dx\,dy|^2}{\int |f|^2\,dx\,dy} \tag{336}$$

where f is the distribution of the dipole amplitudes over the x-y plane, in this case the open face of the waveguide. As we recall, the distribution of electric field in the $TE_{1,0}$ mode is uniform along the Y axis and sinusoidal along the X axis. Performing the indicated integrations of Eq. (336) for this distribution, we obtain

$$G_0 = \frac{32ab}{\pi\lambda^2} \tag{337}$$

where a and b are the cross-sectional dimensions of the guide. For a rectangular guide with $ab = \lambda^2/4$ the gain is $8/\pi = 2.5$ or somewhat greater than that of a dipole, as previously stated.

Horn Radiators.—Higher gains may be obtained from waveguide radiators by widening the aperture so that it extends over a larger number of square wavelengths. This is accomplished by

attaching a flare or horn to the waveguide, so designed as to transform the transverse wave at the end of the guide to a similar transverse wave at the mouth of the horn, without exciting higher modes. These requirements restrict the flare angle (Fig. 156) of the horn. Thus, if high gain is required, the necessary wide mouth is obtained only at the expense of a long horn structure.

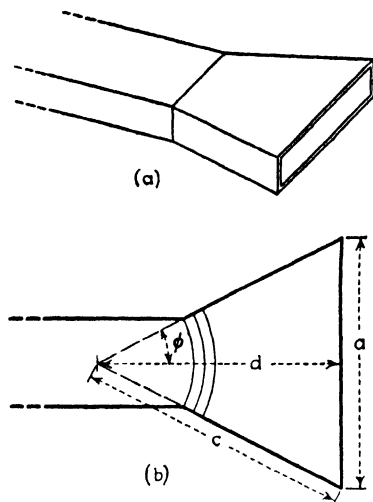



FIG 156.—Horn radiator.

The gain of a horn radiator is the same as that of the open waveguide, Eq. (337), provided that the wavefront at the horn mouth may be considered plane. As shown in Fig. 156, the wave tends to assume a spherical shape as it enters the horn. By restricting the dimension c so that it is but little larger than the dimension d, the portion of the spherical wave appearing at the horn mouth is limited to a small solid angle and hence may be considered essentially plane. This condition is met if the difference between c and d does not exceed a quarter wavelength:

$$c - d = \frac{\lambda}{4} \tag{338}$$

The quantity $c - d$ may be expressed approximately as

$$c - d = \frac{a^2}{8c} \tag{339}$$

where a is the width of the horn mouth. Then the sine of the flare angle is limited to

$$\sin \phi = \frac{a}{2c} = \frac{\lambda}{a} \tag{340}$$

Thus the quantity $a \sin \phi$ should be no greater than the operating wavelength. Equation (340) also implies that the beamwidth λ/a, expressed in radians, must not be smaller than the sine of the flare angle.

These relationships may be used to design sectoral horns, and they apply also to pyramidal horns (Fig. 157) when each aspect of the pyramid is designed as a sectoral unit. Conical horns serve to terminate circular waveguides in the same manner, and the same general geometrical relationships govern their design.

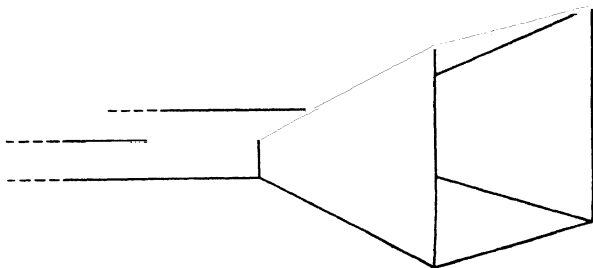



FIG. 157.—Pyramidal horn radiator.

90. Directive Properties of Dipole Arrays.—The dipoles and linear radiators considered in the preceding paragraphs, as well as the simple waveguide radiator, are not highly directive, and consequently they display low values of gain. Our discussion of the horn radiator has indicated that higher gain can be obtained by exciting an area, which is large compared with the wavelength squared, with in-phase currents. Such an area may take the form of flat assembly of in-phase radiating elements, or it may be a virtual surface, such as the opening of a horn radiator or the mouth of a paraboloid reflector.

The properties of such radiating areas may be deduced by considering them to be assemblies, or *arrays* of dipole elements. In this section we consider the properties of such dipole arrays. We treat first the continuous array, all points of which are excited and contribute directly to the radiation. The properties deduced for the continuous array apply very closely to the array of discrete elements, as is indicated in Sec. 96.

The dipole, we recall, is represented by current in a very short and thin wire, the current varying sinusoidally with time but remaining uniform along the length of the wire. The continuous dipole array is made up of a continuous sheet of infinitesimal dipoles and hence is often referred to as a "current sheet." The electric and magnetic potentials previously derived for the dipole, Eqs. (301) and (302), are computed for the array by integrating over the area of the array and by taking into

account variations in current density within the area. When the potentials have thus been computed, the electric and magnetic fields are found, as in the case of the single dipole, by differentiation in accordance with Eqs. (299) and (300).

When this process is carried out in the general case, it is found that the polar distribution of the electric field (the polar diagram) is related to the current distribution in the array by the Fourier integral, the familiar relation we have used in the study of pulse transmission in Chap. II.

To show the application of the Fourier integral, we choose the dipole array illustrated in Fig. 158. The array lies in the $x - y$ plane. The current density varies along the x axis as the function $f(x)$ but is constant along the y axis. The width of the array is a, and the length along the y axis is assumed to be very large. It is further assumed that all the dipole currents are in time phase and that the points of observation are at a great distance, well beyond the induction field. We are required to find the polar diagram of the electric field, that is, its variation with the angle ϕ in the x-z plane.

FIG. 158.—Coordinates for computing the polar diagram of an array of elemental dipoles (current sheet)

Under these conditions, the magnetic potential **A** at a distance r and angle ϕ is given by Eq. (302) as

$$\mathbf{A} = \frac{1}{c} \int_{-a/2}^{+a/2} \frac{f(x)}{r} \epsilon^{j(\omega t - 2\pi r/\lambda \pm 2\pi x \sin \phi/\lambda)} \, dx \tag{341}$$

Since the distance r is very great, r does not vary as the integration is carried across the array from $x = -a/2$ to $x = +a/2$. Moreover we may omit consideration of the time variation and concentrate on the variation of potential with the angle ϕ which is given by the expression

$$F(\phi) = \int_{-a/2}^{+a/2} f(x) \epsilon^{j2\pi x \sin \phi/\lambda} \, dx \tag{342}$$

Now if we confine our attention to small values of ϕ, we may replace sin ϕ by ϕ and obtain the approximate expression

$$F(\phi) = \int_{-a/2}^{+a/2} f(x)\epsilon^{j2\pi x\phi/\lambda}\, dx \tag{343}$$

This is a form of the Fourier integral, as will be recognized by comparison with Eq. (62), Chap. II. Thus we find that the variation of the magnetic potential with the polar angle ϕ can be found from the Fourier spectrum of the current density distribution $f(x)$.

Finally we can identify the variation in electric field with the same expression. This is true because the electric field has but one component, E_y, parallel to the dipoles, and is constant along this direction. Hence, in Eq. (300), $\nabla V = 0$ and

$$E = -\frac{1}{c}\frac{\delta \mathbf{A}}{\delta t} \tag{344}$$

When we take the partial derivative of Eq. (341) with respect to time, only the factor $\epsilon^{j\omega t}$ is affected, and the spatial distributions in r and ϕ remain as multiplying constants. Hence the distribution of the magnetic potential with angle ϕ given by $F(\phi)$ in Eq. (343) applies equally well to the electric field. We find, therefore, that we can plot the variation of field strength against ϕ (the polar diagram) merely by solving Eq. (343). Evidently the polar distribution depends wholly upon the distribution of the current amplitude across the array, given by $f(x)$. For each variation of current $f(x)$ we find a corresponding polar diagram.

91. Directivity of Uniform-current Dipole Arrays.—There are several variations of particular interest. The first we shall examine is the case of uniform current distribution across the array, that is, $f(x)$ is a constant, say unity. Then the Fourier integral becomes

$$F(\phi) = \int_{-a/2}^{+a/2} \epsilon^{j2\pi x\phi/\lambda}\, dx \tag{345}$$

$$= \frac{\sin \pi a\phi/\lambda}{\pi a\phi/\lambda} \tag{346}$$

This is recognized as the $(\sin x)/x$ form that represents the frequency spectrum of a transient rectangular pulse, the uniform amplitude of the pulse corresponding to the uniform current distribution in the array. For a given width of array a and wavelength λ, we can plot the variation as a function of ϕ. This

is shown in rectangular coordinates in Fig. 159*a* and in polar coordinates in 159*b*. The similarity of the rectangular form to Fig. 70, Chap. II, is apparent.

A very great deal of information concerning directive patterns is implicit in these diagrams. First we note that the main lobe of the polar diagram is contained between the first zeros on the $(\sin x)/x$ curve, at the angular values $\phi = -\lambda/a$ and $\phi = +\lambda/a$. Thus the total beam width of the main lobe of signal strength is $2\lambda/a$.

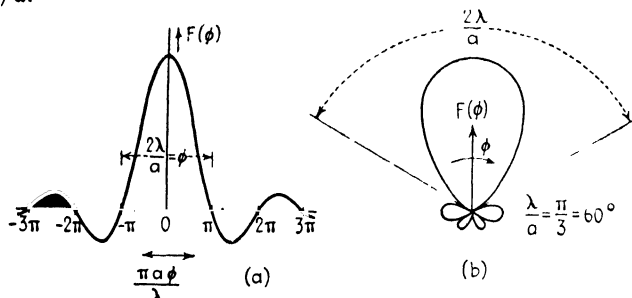



FIG. 159.—Angular distribution of electric field radiated by an array of width *a*, plotted in rectangular (*a*) and polar coordinates (*b*). Compare with Fig. 70, page 103.

The main lobe of *power density*, the quantity referred to in previous discussions, is proportional to the square of the field strength, that is, to

$$F^2(\phi) = \frac{\sin^2 \pi a\phi/\lambda}{(\pi a\phi/\lambda)^2} \tag{347}$$

shown in rectangular and polar form in Fig. 160. The first zeros of this curve occur at the same angular value as the electric field and the total beam width is $2\lambda/a$. Ordinarily, however, we define beam width at the 3 db points, that is, half way down the power distribution curve, as shown in Fig. 160*b*. The beam width at the 3 db points is seen to be

$$b = \frac{\lambda}{a} \quad \text{radians} \tag{348}$$

This expression has already been used in our early discussion of the radar equation (Chap. I, Eq. (14)). Equation (348) expresses a very well-known fact: the main lobe of a dipole array (in this case continuous in form) becomes narrower as the width of the array increases relative to the wavelength. More

specifically, the beam width in radians is equal to the angle subtended by the wavelength at a distance equal to the width of the array.

The distribution of electric field outside the main lobe, given by the portions of the $(\sin x)/x$ curve outside the first zeros, comprises the minor lobes in the radiation pattern, generally referred to as "side lobes." Side lobes are a source of considerable difficulty in practical radar technology. The first aspect is the evident fact that the side lobes remove power from the main beam. A second and more serious problem is

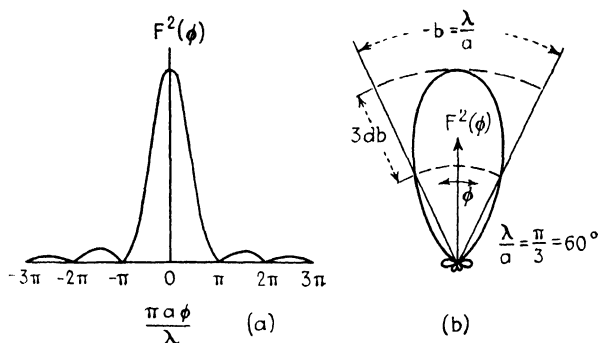



FIG. 160.—Angular distribution of radiated power, formed by "squaring" Fig. 159, in rectangular (*a*) and polar coordinates (*b*).

the possibility of indicating false bearings. Thus a large near-by target centered in a side lobe may return an echo comparable in strength to that from a small or distant target in the main lobe, and confusion may result. A third difficulty, closely allied to the second, is the introduction of ground clutter or permanent echoes through side lobes when the main beam is in the clear. Consequently much attention has been paid to the reduction of side lobes in radar radiators.

In the case here considered, the magnitude of the first side lobe in the power distribution (Fig. 160) is approximately 5 per cent of the magnitude of the main lobe, whereas reduction of side lobes to 1 per cent or less is generally desired, and is usually achieved in practice. It is evident, then, that a uniform current distribution is not desired in practical arrays. Rather, some other distribution must be found that displays smaller side lobes.

It should be noted that the ratio of the lobe magnitudes does not change with the beam width, but only with the current distribution in the array. Although a narrow beam provides room for a larger number of side lobes than does a broad beam, lobes beyond the second may generally be disregarded. Another aspect of the uniform current dipole array is the fact that the beam produced by it is the narrowest that can be achieved for a given ratio of width to wavelength. Thus when we change the distribution of current to obtain smaller side lobes we must necessarily produce a wider beam. Generally this increase in width is small and can be tolerated, or it may be compensated by using a slightly wider array.

The gain of the uniform-current dipole array is defined, as in the case of the isolated dipole, as the ratio of the Poynting vector (proportional to the square of the field strength) along a particular angle to the average value of the Poynting vector over all angles. Thus to arrive at the general expression for gain we take $F^2(\phi)$ at the particular angle of interest and divide by the average value $\overline{F^2(\phi)}$. The maximum gain, at $\phi = 0$, is then

$$G_0 = \frac{1}{\overline{F^2(\phi)}} = \frac{1}{1/2\pi \int_0^{2\pi} F^2(\phi)\, d\phi} \tag{349}$$

In the case of the uniform current distribution here considered, the maximum gain, relative to an isotropic radiator, is

$$G_0 = \frac{\pi a}{\lambda} = \frac{\pi}{b} \tag{350}$$

Thus the maximum gain is equal to the ratio of the half circle to the beam width b between half-power points.

Thus far we have treated the polar diagram of a semi-infinite array, that is, one of definite width in one dimension but of infinite length in the other dimension. In this way we have connected the polar diagram in a particular plane with the width of the array measured in the same plane.

When we consider a finite array, with limited dimensions in both width and height, we apply the same process by double integration over the dimension a in the x dimension and b in the y dimension, as shown in Fig. 161. We must also introduce another polar angle θ to measure the polar distribution in a plane

at right angles to the ϕ plane. Then, by a process exactly analogous to that of Eqs. (341) to (345) we arrive at the polar distribution function of the rectangular uniform-current array:

$$F(\phi,\theta) = \sin\theta \frac{\sin[(\pi b/\lambda)\cos\theta]}{(\pi b/\lambda)\cos\theta} \frac{\sin[(\pi a/\lambda)\sin\phi\sin\theta]}{(\pi a/\lambda)\sin\phi\sin\theta} \quad (351)$$

A three-dimensional representation of this polar function is shown in Fig. 162. We note that we now have a product of

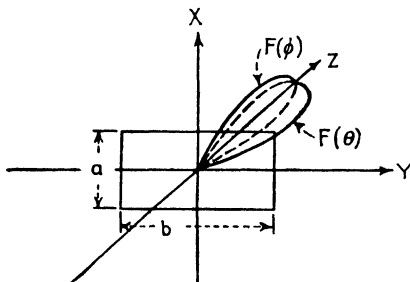



FIG. 161.—Coordinates for angular distribution of an electric field from a rectangular array of height (a) and width (b).

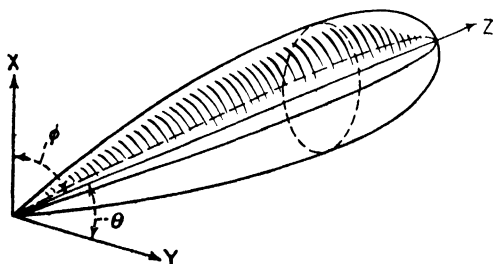



FIG. 162.—Angular coordinates (ϕ and θ) of a conical beam.

two $(\sin x)/x$ functions, and that there is no radiation in the plane of the array along the b dimension ($\theta = 0$), which is to be expected since none of the individual dipoles radiates along its length.

To find the gain of the rectangular array, we find the average square value of Eq. (351)

$$\overline{F^2(\phi,\theta)} = \frac{\lambda^2}{4\pi ab} \quad (352)$$

Then by Eq. (349), the maximum gain, relative to an isotropic radiator, is

$$G_0 = \frac{4\pi ab}{\lambda^2} \quad (353)$$

Thus we find that, in the case of a rectangular array, the gain is determined by the area of the rectangle ab expressed in square wavelengths. Moreover, the gain is inversely proportional to the ratio of the solid angle λ^2/ab subtended by the beam to the whole solid angle of a sphere, 4π steradians.

It will be noted that, in the gain equations for the semi-infinite array, Eq. (350), and the rectangular array, Eq. (353), zero gain is indicated for very narrow and very shallow arrays. But since the limit of both arrays is a single dipole, and since the gain of a single dipole is 1.5, both equations are evidently in error for small gains. Both are sufficiently accurate for use in practical cases, that is, when the gain exceeds 10. Gains above one hundred are possible in rectangular arrays, provided that the linear dimensions exceed three times the wavelength. Such a ratio is quite practical when the wavelength is measured in centimeters (microwave region) but less so when the waves exceed a meter in length.

By similar methods (involving Bessel functions) the polar distribution and gain of a uniform current array of circular shape and radius r may be computed. The gain equation is

$$G_0 = \frac{4\pi(\pi r^2)}{\lambda^2} \tag{354}$$

This expression is similar in form to Eq. (353) for the rectangular array, and it provides identical gain for the same area ($ab = \pi r^2$). It is subject to the same error when applied to arrays of low gain.

92. Single-sided Radiation from Arrays.—In the foregoing computations of gain, we have taken the average radii of the polar diagram as though radiation issued from but one side of the array, which is the case in practical radiators. But since a current sheet is symmetrical with the positive and negative directions of propagation, it must necessarily radiate two identical patterns in opposite directions. When this fact is taken into account, the gains stated in Eqs. (353) and (354) are halved.

In practice, such two-sided radiation would be wasteful and harmful, and thus means are taken to prevent it. Either of two means is generally employed: a solid metal reflector through which the backward radiation cannot penetrate, or a parasitic reflector array, spaced one-quarter wavelength behind the

excited array (the radiator proper). The action of these devices depends on the formation of two current sheets, whose separation and phase difference are such as to cancel the radiation in the backward direction and reinforce it in the forward direction.

The situation is most clearly seen in the case of two current sheets separated a quarter wavelength (Fig. 163). Consider two individual dipoles opposite each other along the axis of the main lobe. The two dipoles are considered to have equal current amplitudes. The left-hand dipole is the excited dipole, the right-hand one being excited parasitically from the first. The time phase of the parasitic dipole is one-quarter cycle later than that of the primary dipole. Under these circumstances,

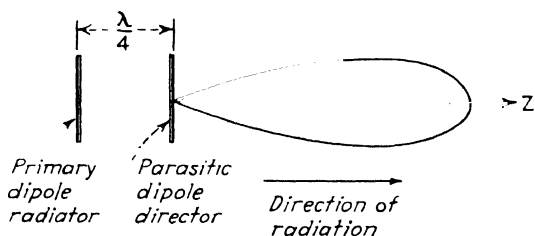



FIG. 163.—One-sided radiation from two dipole elements excited in quadrature.

the radiation from the excited dipole reaches the parasitic dipole in phase and remains in phase as both proceed outward along the positive z axis. Hence the radiation along this axis is twice that from the excited dipole alone. In the opposite direction, however, a reversed situation occurs. The energy from the parasitic dipole, already a quarter cycle behind in time initially, lags a half cycle on reaching the excited dipole. There the two fields just cancel and remain canceled as they proceed outward along the $-z$ direction. Hence, no field is propagated in the direction from parasite to primary, and double field is propagated in the opposite direction. This is the fundamental action of a director type of array, in which the radiation is reinforced in the direction of the parasitic element. The parasitic reflector is similar, but the time phase of the parasitic currents is arranged, by virtue of tuning, to be a quarter cycle *in advance* of the primary current, so that the outward radiation proceeds in the direction from parasite to primary.

This simple demonstration of one-sided arrays is limited to the z direction (plus and minus). At directions off the z axis

the cancellation and reinforcement is less complete. Detailed computation of the field as a function of the polar angle ϕ shows it to vary as a cardioid, shown in Fig. 164.

When an array of such cardioid patterns is assembled, the net field is substantially the same as that of the simple current sheet, except that no radiation occurs in one direction, and double values of field are found in the other. Hence the gain equations previously derived on the basis of one-sided propagation apply without change to one-sided arrays.

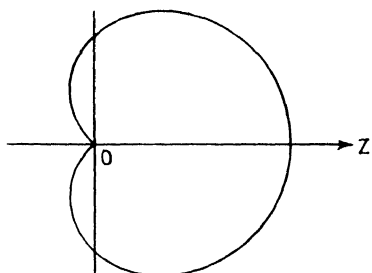



FIG. 164.—Cardioid radiation pattern from dipole and reflector.

93. Reduction of Side Lobes by Current Distribution.[1]—Among the directional properties of current sheets, we must consider current distributions that minimize radiation in side lobes. This is an important practical matter, as previously indicated. The clue to reduction of side lobes is found in the Fourier integral relationship between current distribution and polar distribution of field. Our experience with the Fourier integral in Chap. II indicates that a wide variety of pulse shapes possess spectra of the undulating type, but that the undulations become less prominent as the pulse shape departs from the rectangular and becomes more rounded. The corresponding case in the dipole array is a current distribution which does not cease abruptly at the edges of the array, but which gradually approaches zero. An example is the current distribution according to the Gauss error curve. The Fourier integral of the error function is another error function. Since the error function itself has no undulations, we conclude that the polar diagram of such a current distribution is free of side lobes. Suppose we take, as the distribution of current along the x and y axes of a rectangular array, the following form:

$$f(x,y) = \frac{k}{\pi} \epsilon^{-k(x^2+y^2)} \tag{355}$$

[1] DOLPH, C. L., A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level, *Proc. I.R.E.* **34,** 335 (June, 1946).

a typical two-dimensional error function. Then the polar diagram $F(\phi,\theta)$ turns out to be

$$F(\phi,\theta) = \sin\theta\epsilon^{-(m^2/4k)(\cos^2\theta+\sin^2\theta\sin^2\phi)} \qquad (356)$$

which is a function which approaches zero indefinitely as the angles ϕ and θ become larger up to $\pi/2$. Thus there are no discrete sidelobes. But there is some small radiation, by the same token, at angles far removed from the forward direction The radiation may be made substantially zero outside any limiting values of ϕ and θ by suitable choice of the constant k that describes the current distribution across the face of the array. When we adjust the value of k we find that the gain decreases as the side radiation is reduced. Thus we find that a compromise must always be made between high gain and low side radiation (whether in lobes or not). Both requirements cannot be met simultaneously.

Other forms of gradually diminishing current, such as that described by a cosine raised to a power, give results similar to the error function. The curve of $\cos^4$, for example, has half the gain of the uniform current distribution but side lobes that disappear faster by a factor containing the fourth power of the polar angle.

94. Lobe Shifting by Phase Relationships.—The preceding derivations of gain and polar distribution have been based on oscillations that are assumed to be in the same time phase, at all points in the current sheet. This condition is realized in many practical forms of array, including the waveguide horn, paraboloid reflector, and arrays of discrete halfwave elements. The equiphase relationship produces a symmetrical lobe of radiation, directed along the normal to the radiating surface. When shifts of the beam axis are required, for scanning, lobe switching, or other purposes, it may prove impractical to move the array or reflector at the required speed. Shifting of the lobe axis may be achieved electrically by introducing phase changes along the surface of the current sheet. Such phase changes may be produced by introducing different physical separations between the radiator proper and various parts of the associated reflector or by introducing electrical delays in the feeds to various portions of an array of discrete elements.

The general effect of phase changes in the current sheet may

be seen from the Fourier integral, Eq. (343). Suppose that the distribution of the current elements, represented by $f(x)$, has a constant amplitude I across the array but a linear phase shift of α radians from one edge to the other (Fig. 165). Then the distribution of current elements can be written

$$f(x) = I\epsilon^{-j\alpha x/a} \tag{357}$$

When this expression is substituted in Eq. (343), the exponent $-j\alpha x/a$ adds to the exponent $j2\pi x\phi/\lambda$, and the total exponent can be written

$$\frac{j2\pi x}{\lambda}\left(\phi - \frac{\alpha\lambda}{2\pi a}\right) \tag{358}$$

The polar distribution of the field is then given by Eq. (347) except that ϕ in that equation is now replaced by $\phi - \alpha\lambda/2\pi a$. This corresponds to a shift in the direction of the lobe of $\alpha\lambda/2\pi a$ radians relative to its position for the same current sheet with equiphase oscillations. It is evident that the shift of the lobe axis may be increased by increasing the phase shift α, or the wavelength λ, relative to the array width a. It should be remarked, however, that the above relationship applies only for small shifts, since the use of ϕ rather than $\sin \phi$ in Eq. (343) is based on small angles.

FIG. 165.—Amplitude and phase angle of array excitation that produces an angular shift in the beam axis.

If the phase shift across the face of the dipole array is not linear, the shift in direction occurs but is no longer given by Eq. (358) and is accompanied by a change in the relative amplitudes along the different polar directions. Such distortion of the lobe shape may usually be avoided since the linear phase shift is not difficult to achieve. Several practical radiators employing phase shift for lobe switching or scanning are described in Chap. VII.

95. Physical Forms of Dipole Arrays.—The current sheets used as the basis of the foregoing analysis are, of course, theoretical in nature and are never exactly reproduced in practice. Among the physical forms of interest are the "mattress" array of discrete halfwave radiators and reflectors and the paraboloidal reflector

illuminated by a dipole or waveguide radiator. The intensity of current in the aperture plane of such arrays is never a simple analytic function and hence the polar diagrams and gains of such radiators are difficult to compute accurately. However, measurements indicate that the directivity properties can be predicted, within a few percent in angle and amplitude, by the use of the relationships previously derived. Effort is expended to produce a gradual diminution of the current density toward the edges of the array, in the interest of small side lobes, and to avoid loss of energy by resistive heating in poor conducting

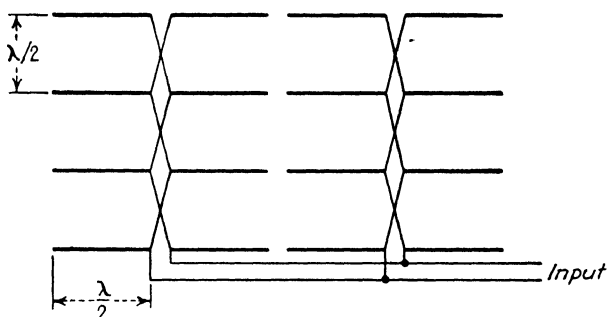



FIG. 166.—In-phase excitation of dipole elements.

surfaces. But when these measures have been taken, the laws of geometric optics leave little freedom of design. Sharp beams can be obtained only at the expense of areas large with respect to the square wavelength, and no amount of effort can alter that fact.

96. Arrays of Discrete Half-wave Elements.—The mattress array (Figs. 166 and 167) is in many respects the closest approach to the theoretical current sheet, because the separate elements can be fed with currents of equal amplitude in the same time phase. The essential departure from the current sheet is the discrete nature of the current elements. Each element of a mattress may be thought of as a compact gathering of current dipoles that might otherwise be spread uniformly over the area of the array. Evidently such gathering of the current elements does not affect the total energy radiated, but it may affect the polar distribution of field if the discrete elements are too far separated. The rule is that the elements should not be farther than a wavelength apart. Usually a half-wavelength separation

is employed, although separations up to 90 per cent of a wavelength serve adequately and may, in fact, give smaller side lobes.

The discrete array may be treated quantitatively by adding vectorially the fields from the several halfwave elements at points along each direction in the polar plane. The polar distribution than can be shown to be

$$F(\phi) = \sum^{n} \epsilon^{jn2\pi d/\lambda} \sin \frac{\phi}{\lambda} \tag{359}$$

where n is the number of elements and d/λ is the spacing between elements in wavelengths. This summation for large values of

FIG 167.—Typical "mattress" dipole arrays for transmission and reception (SCR-268 radar, 205 mc)

n and small values of d/λ approaches the $(\sin x)/x$ form of the continuous current sheet. So long as d/λ remains below the value 1, the difference between the discrete and continuous formulations may be neglected, particularly with reference to the uncertainties in physical dimensions and phases of practical arrays.

Since this is the case, it is permissible to compute the gain of a mattress in terms of the rectangular equation previously derived, Eq. (353), substituting an empirical constant K for the theoretical constant 4π.

$$G_0 = \frac{Kab}{\lambda^2} \tag{360}$$

where a and b are the dimensions of the array. Practical values of the constant K vary from 10 to 12, against $4\pi = 12.6$, the theoretical value. This indicates that the power efficiencies of properly adjusted mattress arrays vary from 80 to 95 per cent. The beam widths (between half-power points) of mattress arrays are given by the theoretical equation Eq. (348) in each dimension, with empirical factors taken into account.

$$B_a = \frac{\lambda}{a}\sqrt{\frac{4\pi}{K}} \quad \text{radians} \tag{361}$$

where B_a is the beam width associated with the a dimension and K is the constant in Eq. (360). It should be noted that, in general, the beam width increases with the square root of the loss in gain.

Another expression for gain may be found by combining Eqs. (360) and (361)

$$G_0 = \frac{4\pi}{B_a B_b} \tag{362}$$

where B_a is the beam width in the a dimension and B_b the beam width in the b dimension, both in radians. This expression indicates that the gain is equal to the solid angle of a sphere (4π steradians) divided by the solid angle of the beam between half-power points ($B_a B_b$ steradians). This is a fundamental proposition that applies to all directive radiators. In fact, Eq. (362) may be taken as the definition of gain and it will then be found that our definition in terms of the maximum and average Poynting vectors is a necessary consequence.

97. Paraboloidal Reflectors.—The paraboloidal reflector is a widely used device for producing directional radiation from a substantially nondirective source, such as a single dipole. Geometrically, the paraboloid is a surface of revolution formed by rotating a parabola about its axis. The points on a parabola, denoted by x and y in Fig. 168, are defined by the equation

$$x = \frac{y^2}{4p} \tag{363}$$

where p is the distance measured along the axis from a point on the axis, known as the "focus," to the curve. As a consequence

of this relation, a ray of radiation, proceeding from the focus and reflected by the paraboloid, proceeds after reflection along a path parallel to the axis.

The action of a paraboloid excited by r-f energy is evident when we consider a plane ("face place") at right angles to the axis. All reflected paths from the focus to this plane are of the same length, that is, a given path *abc* (Fig. 168) has the same length as any other path *ade*. Hence radiation leaving the focus in phase arrives at the plane in phase and constitutes a current sheet of in-phase dipoles. The intensity of illumination is not constant across this plane. Rather the intensity drops off gradually toward the edges of the paraboloid, a condition favorable to the reduction of side lobes. This nonuniformity has the effect of reducing the maximum gain of the paraboloid, relative to that of a uniform current sheet of the same area. The reduction in gain and the effect on the side lobes depend greatly on the form of radiating source and on the shape of auxiliary shields which are used to conserve energy which would otherwise be radiated directly from the source in the forward direction.

FIG. 168.—Geometry of the paraboloidal reflector.

The distance from the face plane to the focus is usually taken as zero, that is, the face plane passes through the focus. The choice of this dimension depends on the distribution of energy from the feed. If the feed is substantially a point source, a shallow paraboloid (Fig. 169*a*) wastes energy since the reflector intercepts only a small portion of the total energy radiated. On the other hand in a deep paraboloid (Fig. 169*b*) the area of the face plane (which determines the beam width) is small compared with the area of the paraboloid; hence the structure is unwieldy. The gain is a maximum somewhere between these two cases. If the point source is hemispherical (transmits through a half-sphere in the direction of the paraboloid), the gain is a maximum when the face plane passes through the focus. The gain is then about 96 per cent of the gain of a uniform current sheet of the same area.

In practice, since a hemispherical point radiator cannot be achieved, the gain of practical feeds and associated paraboloids is generally less, by a substantial amount, than the equivalent current sheet. If the feed is an open dipole at the focus without shield (Fig. 169*c*), the gain is about 35 per cent of the theoretical maximum. A dipole with a simple reflector shield (Fig. 169*d*)

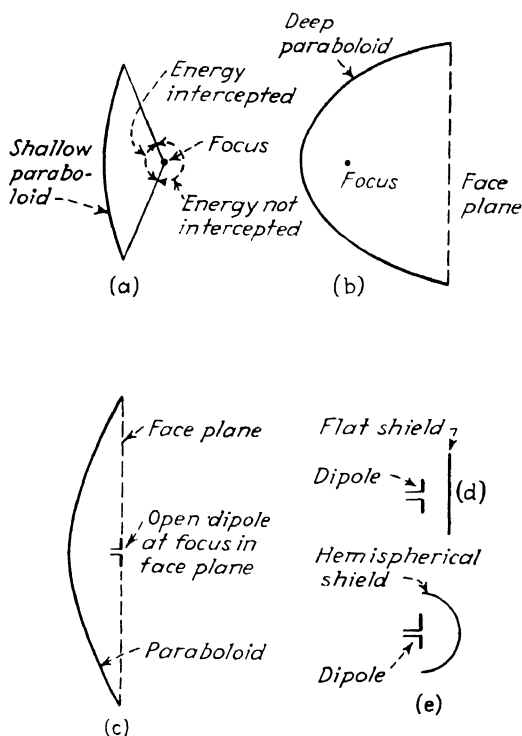



FIG. 169.—Paraboloidal reflectors and feeds.

has a gain about 45 per cent of the maximum. A dipole fitted with a hemispherical shell approaches the hemispherical point source, and the gain is about 60 per cent of the maximum.

It will be noted that the gain of the paraboloid is generally lower than that of a mattress of the same area expressed in square wavelengths. The choice between paraboloid and mattress is usually made not on the basis of relative gain, however, but on the basis of mechanical and electrical convenience. The latter consideration dictates paraboloidal surfaces for the

higher frequencies (shf or microwave region) and mattress arrays for the lower frequencies.

Exact calculation of paraboloid polar diagrams and gains is seldom possible because of uncertainties in the distribution of the energy in the feed. The gain may be computed for simple assumed distributions, however, by reference to the basic gain equation

$$G_0 = \frac{4\pi}{\lambda^2} \frac{|\int f\, dx\, dy|^2}{\int |f^2|\, dx\, dy} \tag{336}$$

Taking the polar angle θ in the plane of the parabola about the focus as an origin, we can write the equation of the parabola as

$$r = \frac{2p(1 - \cos\theta)}{\sin^2\theta} \tag{364}$$

and the element of area of the face plane as

$$dx\, dy = da = r^2 \sin\theta\, d\theta\, d\phi \tag{365}$$

where ϕ is the polar angle in the face plane. Finally we take the distribution of the electric field between source and face plane as

$$f = \frac{g(\theta,\phi)}{r} \tag{366}$$

which indicates that the field falls off as the radial distance r from the source. Then, substituting in Eq. (336), we obtain

$$G_0 = \frac{16\pi p^2}{\lambda^2} \frac{|\int g(1 - \cos\theta)/\sin\theta\, d\theta\, d\phi|^2}{\int |g|^2 \sin\theta\, d\theta\, d\phi} \tag{367}$$

When a unit hemispherical point source at the focus is assumed as the feed, we take $g = 1$ for values of θ between 0 and $\pi/2$, and $g = 0$ for all other values. Evaluation of the integral then yields, very closely,

$$G_0 = \frac{16\pi p^2}{\lambda^2} 2\pi(\log_e 2)^2 \tag{368}$$

This value of gain applies for any value of the focal distance p. The maximum gain is found when the focus lies in the face plane and $p = a/2$, where a is the radius of the face plane. Then

$$G_0 = 0.96 \frac{4\pi(\pi a^2)}{\lambda^2} \tag{369}$$

Comparison with Eq. (354) shows that the gain is 96 per cent of that of an equivalent uniform current sheet. A similar computation may be based on cosinusoidal variation of g with θ and ϕ, representing a single unshielded dipole as the feed.

The empirical approach to the gain of a paraboloid radiator is parallel to that for the mattress, Eqs. (360), (361), and (362). The applicable value of the constant K depends on the nature of the feed and its shield. A well-shielded dipole feed may display a value of K between 5 and 7.5, corresponding to 40 and 60 per cent of the gains of a uniform current sheet.

Waveguide radiators are also used as feeds for paraboloids. Since the directional properties and gain of a matched open-ended waveguide radiator are approximately the same as that of a dipole and reflector, the same value of the constant K is generally applicable to both types.

It has been found experimentally that an advantage may be obtained by placing the feed radiator a slight distance away from the focus of the paraboloid. This displacement causes the equivalent dipoles in the face plane to depart from the equiphase condition, symmetrically about the axis. The theoretical effect of the nonuniform phase distribution is a loss of gain and a secondary effect on the magnitude of the side lobes. Maximum gain is obtained when the distance from feed to reflector, measured along the axis, is a whole number of half wavelengths. Then the back radiation from the feed (which is always present in some degree) adds in phase with the reflected radiation along the axis, and a maximum signal is obtained. If the focus is not a whole number of half wavelengths from the reflector, a shift of the feed from the focus may increase the gain.

98. Paraboloidal Sections.—The paraboloid of circular shape, just discussed, produces a beam of radiation that is approximately symmetrical about its axis. A slight dissymmetry occurs due to the nonuniform distribution of the radiation from the dipole feed, but this effect is small. When it is desired to produce a beam appreciably wider in one cross-sectional dimension than in the other, a paraboloidal section that has an approximately rectangular outline may be used. A typical example is shown in Fig. 170. The gain of such a structure is given, approximately, by the expression for a rectangular array of half-

wave elements, Eq. (360). The applicable value of the constant K, depending on the orientation of the dipole feed, varies from about 4 to 7. The beam widths in each of the two dimensions are given by Eq. (361), using the same values of K.

A paraboloid with a rectangular (or other noncircular) face has the disadvantage that the radiation from the feed cannot always be confined so that it falls only on the reflector surface. Thus, if a hemispherical point source is assumed at the focus, full illumination may be achieved over the long dimension when the face plane passes through the focus, but the missing portions of the shorter dimension allow radiation from the feed to pass in the backward direction. This is not only a loss from the main beam, but constitutes a source of undesirable back radiation.

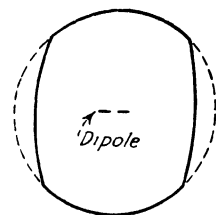



FIG. 170.—Paraboloidal section.

To obviate these difficulties, it is usual to employ a parabolic cylinder, that is, a surface which has a parabolic cross section in one dimension only (Fig. 171). Such a reflector is directive in one plane only, that containing the parabolic section. If such a partial paraboloid is fed from substantially a point source,

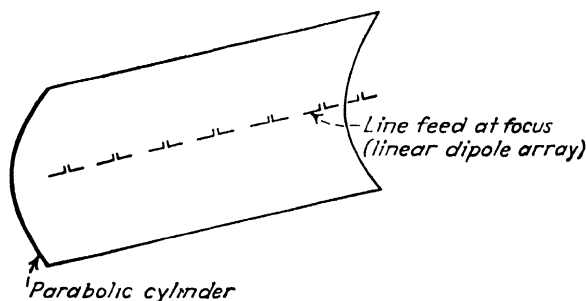



FIG. 171.—Paraboidal cylinder and linear dipole array as feed

the radiation in the parabolic plane is confined to a beam width given by Eq. (361), whereas the beam width is that of the dipole source in the other dimension and the resulting beam is fan shaped.

More commonly the feed is extended over a substantial portion of the line focus of the reflector. Usually a linear array of dipoles,

a thin waveguide radiator, or a slit in the side of a waveguide is employed. This confines the radiation in the plane passing through the line focus, in accordance with Eq. (361). By suitably proportioning the width in the parabolic section and

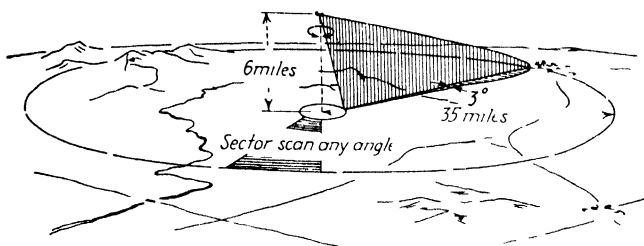



FIG. 172a.—Cosecant-squared distribution for scanning extended areas from an elevated point.

the length of the feed, any desired shape of beam may be obtained, from a thin fan to a symmetrical cone.

Reflectors for Asymmetrical Energy Distributions.—Two paraboloidal surfaces of different focal length may be joined (Fig. 172*b*) to produce a nonuniform distribution of radiated energy with the beam. The need for such a distribution is evident, for example, when an airborne radar scans the surface of the earth forward along the line of flight. The ground directly beneath the aircraft, being closest, returns a strong signal. The surface (Fig. 172*a*) some distance in front of the plane, on the other hand, is often the region of principal interest. Because of the greater distance this region receives and returns a weaker signal. To provide uniformity in the strength of the echoes it is desirable to confine most of the energy to the forward direction and to reduce the energy radiated directly downward. The radiated signal strength should vary inversely as the sine of the nadir angle, or directly as the cosecant of that angle. Hence the

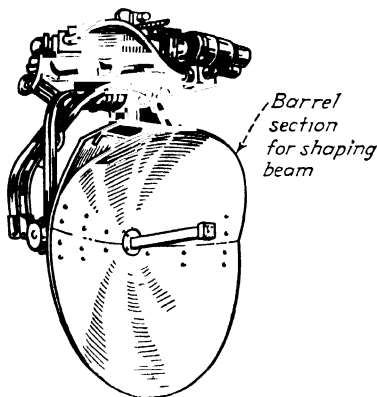



FIG. 172*b*.—Distorted paraboloid for produced cosecant-squared pattern

radiated power distribution must vary as the cosecant squared. Since this relationship applies also to the received echo, the net effect on the signal power delivered to the receiver is to introduce a cosecant raised to the fourth power and thus to counteract the inverse fourth power in the radar equation.

99. Absorptive Properties of Directive Radiators.—The ability of a directive radiator to absorb incident radiation must be considered for receiving purposes. Since all modern radar equipment uses the same antenna for transmission and reception, it is convenient to discuss the radiation and the absorption together, despite the fact that, chronologically speaking, propagation and target reflection occur between the two.

Absorption by directive radiators may be discussed in terms identical to those previously used, with two important additions: polarization of the incident energy and scattering. A transmitting radiator confines itself to radiating energy, the polarization of which is determined by the orientation of the dipole or waveguide source. In reception, on the other hand, the polarization may have been rotated or robbed of its plane properties by the reradiation characteristics of the target. Also a receiving antenna always performs two functions: it absorbs energy and transmits it to the receiving circuits, and it also reradiates or scatters energy. The primary act is absorption, that is, the abstraction of energy from the incident wave. Part of the absorbed energy is utilized in the receiving circuits, and part of it is reradiated. The maximum amount of energy is absorbed when the electrical (circuit) resistance connected to the dipole or waveguide radiator is equal to the radiation resistance. Then the absorbed energy is divided equally between the utilized and the reradiated forms.

The total power absorbed is computed in terms of the power density falling on the antenna and the so-called "absorption cross section" of the antenna, which is defined as

$$\sigma = \frac{P_a}{S_r} \tag{370}$$

where P_a is the total power absorbed and S_r is the power density received. It is evident that the absorption cross section varies with the polarization of the incident wave, with the tuning of the dipole element or elements, and with the direction from

which the incident radiation arrives. Usually we are interested in the maximum absorption cross section, which occurs when the electric vector is polarized parallel to the dipole elements, the dipoles are tuned to resonance, and the direction is normal to the array or face plane of the reflector. When these conditions are satisfied and when the incident wave is essentially plane, we can compute the maximum and average values of σ not only for a single dipole, but also for any array or reflector whose maximum gain G_0 is known.

We consider, first, a single resonant dipole, on which a plane wave whose rms amplitude is E volts per m (peak amplitude $1.4\,E$) is incident. Then the Poynting vector gives the average power density incident as

$$S_r = \frac{c}{8\pi} E^2 \tag{371}$$

Also the average power absorbed may be computed to be

$$P_a = \frac{3\lambda^2 cE^2}{64\pi^2} \tag{372}$$

Then, by Eq. (370), the absorption cross section is

$$\sigma = \frac{3\lambda^2}{8\pi} \tag{373}$$

Thus the absorption cross section of a resonant dipole is equal to the area of a circle whose radius is about 40 per cent of the wavelength, or a square whose side is about 35 per cent of the wavelength.

This is the cross section that applies for parallel polarization. It is also of interest to find the average cross section when the incident radiation has any angle of polarization ϕ relative to the dipole axis. The power absorbed varies with $\sin^2 \phi$, the average value of which is $2/3$. Then

$$\sigma_{av} = \frac{\lambda^2}{4\pi} \tag{374}$$

This is the area of the circle (πr^2) whose circumference ($2\pi r$) is equal to the wavelength of the incident radiation.

The maximum cross section of a directive system whose maximum gain is G_0 relative to a point source is $2/3 G_0$ times as

great as that of a single resonant dipole. Thus

$$\sigma_{\max} = \frac{G_0 \lambda^2}{4\pi} \tag{375}$$

and the average value for random polarization is

$$\sigma_{av} = \frac{G_0 \lambda^2}{6\pi} \tag{376}$$

Reference to the gain equations of various arrays, Eqs. (353) and (354), and substitution in Eq. (375) show that the maximum absorption cross section of such arrays is equal to their physical area, which is a plausible result [cf. Eq. (20) in Chap. I].

100. Formation of Beams by Electromagnetic Lenses.—The paraboloidal reflector has an analogue in optics in the reflectors

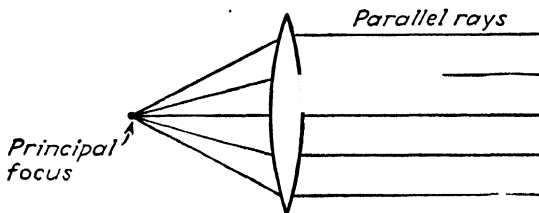



FIG. 173.—Focal action of an optical lens.

that are used to form beams, for example, in automobile headlamps. This analogy suggests another, that is, the electromagnetic lens. The formation of r-f beams by lens structures is comparatively new in the radio art.[1]

The optical action of a lens (Fig. 173) depends on the change of the velocity of the light ray as it passes through the lens element. Thus if a point source of light is located at the focus of a positive spherical (convex) lens, the light energy reaches the center portion of the lens surface first, whereas the light reaching the peripheral region travels a greater distance and is delayed a corresponding amount. The physical action of the lens is to introduce a differential delay such that all rays, having emerged from the far side of the lens, pass a plane at a fixed time relative to their emergence from the source. Since the emerging energy is then located on an equiphase surface, the rays thereafter travel parallel to each other, except for the ines-

[1] KOCK, W. E., Metal-lens Antennas, *Proc. I.R.E.*, **34**, 828 (November, 1946).

capable effects of diffraction at the edges of the lens. Thus the rays from a point source are collimated into a parallel beam. Similarly beams may be formed from line sources by cylindrical lenses.

An exactly analogous procedure may be employed in forming beams of r-f energy from point or line sources. The necessary delay is introduced by passing the beam through an array of open-ended waveguide segments (Fig. 174). The waveguide segments are so proportioned, in cross section, as to be near the cutoff condition at the wavelength in question. Then, as we have seen in Chap. III, the phase velocity during passage through the segments is considerably greater than the velocity in free space. The necessary differential delay between segments is obtained by variation of the length of each segment, that is, by the contour of the lens surface. In this respect the electromagnetic lens is inversely analogous, in geometry, to the optical lens. The surface of the electromagnetic lens array may have a spherical contour, like that of a spherical lens. A concave electromagnetic lens has the same effect as a convex optical lens. When the source is a point source, a cylindrical contour suffices. Special contours, like those of the Fresnel lens, may be introduced for particular purposes.

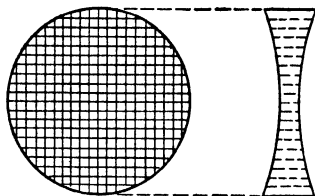



FIG. 174.—Electromagnetic lens composed of waveguide segments. It has the same action as the optical lens shown in Fig. 173.

The advantage of the electromagnetic lens over the paraboloid is the tolerance permissible in construction. At high frequencies, the permissible variation in a paraboloidal surface is of the order of a few hundredths of an inch, which requires careful machining in manufacture and subsequent care in use. The lens, on the other hand, may be distorted by as much as a quarter wavelength (about 1/8 in. at 10,000 megacycles) before adverse effects are produced. The disadvantage is the fact that the segments must be not smaller than a half wavelength on a side, to avoid excessive attenuation of the signal. Thus an adequate mechanical structure is possible only at the higher frequencies.

101. Propagation of Radar Signals.—We turn now to the factors governing the radar signal after it leaves the radiator,

that is, the propagation of the signal to and from the target. We have treated one simple case of radar propagation in Sec. 28, Chap. I, in the derivation of the free-space radar equation. In that case we were concerned simply with the dispersion of the signal through empty space, (1) as it passed from the transmitter, and (2) as it returned from the target. In this discussion, it was pointed out that the maximum range of a radar system, under free-space conditions, varies as the fourth root of (1) the transmitter power, (2) the receiver power sensitivity, (3) the target echo area, and (4) the square of the maximum antenna gain.

We must now extend these conclusions to include the propagation of radar signals when ground or sea reflections are present. Moreover, we cannot content ourselves with the maximum range obtainable along the axis of maximum antenna gain, because the reflected portion of the signal usually starts out from the antenna at an angle below the axis of maximum gain. We must, in short, generalize the free-space radar equation to take account of off-axis radiation and reflections.

In so doing it is convenient to introduce a factor that expresses the departure of the radiated signal from its maximum value as well as the departure of the propagation path from free-space conditions. This factor has, accordingly, been called the "pattern-propagation" factor[1] and is generally given the symbol F.

To examine the nature of the pattern-propagation factor, we recall the free-space radar equation in the form given by Eq. (22), which expresses the maximum free-space range of a radar in terms of the radiator area A.

$$r_{max} = \sqrt[4]{\frac{P_t A^2 \sigma}{P_{rmin} \lambda^2 (4\pi)}} \tag{22}$$

Recalling from Eqs. (353) and (354) that the maximum gain of an antenna is $G_0 = 4\pi A/\lambda^2$, we find the maximum range in terms of the maximum gain of the radiator

$$r_{max} = \sqrt[4]{\frac{P_t G_0^2 \sigma \lambda^2}{P_{rmin} (4\pi)^3}} \tag{377}$$

[1] This terminology was suggested by Kerr and Rubenstein of the Radiation Laboratory, Massachusetts Institute of Technology, in *Rep.* 406 (Sept. 16, 1943).

This is the range in the direction of maximum gain. Suppose now we require the range along some other direction, say along horizontal angle ϕ and vertical angle θ. We denote the particular value of antenna gain along this direction as $G(\phi,\theta)$. The maximum free-space range in this direction is then

$$r_{max}(\phi,\theta) = \sqrt[4]{\frac{P_t G^2(\phi,\theta)\sigma\lambda^2}{P_{r\,min}(4\pi)^3}} \tag{378}$$

We may equally well express the maximum range in a particular direction in terms of the maximum antenna gain and of a factor to convert from maximum antenna gain to gain along the direction ϕ and θ. This factor, squared, is the pattern-propagation factor previously mentioned. Thus

$$G(\phi,\theta) = G_0 F^2(\phi,\theta) \tag{379}$$

Substituting Eqs. (377) and (379) in Eq. (378), we obtain the maximum range along ϕ and θ.

$$r_{max}(\phi,\theta) = r_{max} F(\phi,\theta) \tag{380}$$

Thus the pattern-propagation factor F, a function of the angles ϕ and θ, expresses the proportionality between the maximum range along the axis of maximum gain, and the maximum range along the direction specified by ϕ and θ.

Moreover, since by Eq. (379),

$$F(\phi,\theta) = \sqrt{\frac{G(\phi,\theta)}{G_0}} = \frac{E(\phi,\theta)}{E_0} \tag{381}$$

we find that the pattern-propagation factor is the ratio of the electric field strength $E(\phi,\theta)$ in the direction ϕ,θ to the electric field strength E_0 along the direction of maximum gain.

Thus far we have defined the factor F as it applies to off-axis propagation in free space. The same concept applies when reflections are present, since the effect of reflections is to increase or decrease the electric field strength in particular directions. When the direct and reflected waves reinforce each other at a particular point, the value of F along the direction ϕ,θ to that point is increased. If the waves tend to cancel, the value of F is decreased. The effect of reflections is thus seen to change the *value* of F, but not to change its *effect* on the maximum range.

When reflections are present, we compute F as a ratio of

electric field strengths, in accordance with Eq. (381), and apply the value to Eq. (380) to obtain the maximum range in the presence of reflections, relative to the maximum range under free-space conditions, given by Eq. (377). We thus reduce the determination of maximum range in a radar system to two factors: the free-space maximum range, and the pattern-propagation factor that reduces or increases the range relative to the free-space value. Since the free-space range has already been discussed at length in Chap. I, we confine our attention here to the factor F and compute its value under particular conditions of interest.

102. Propagation over Flat Reflecting Earth.—The determination of the pattern-propagation factor in actual cases is com-

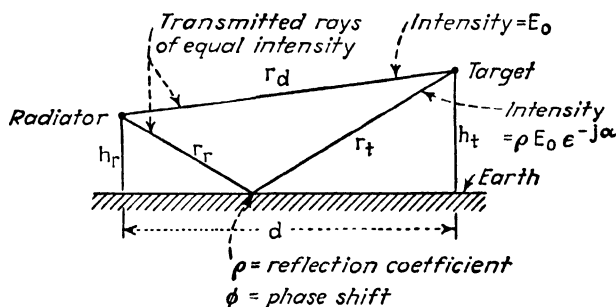



FIG. 175.—Geometry of radio-wave propagation over flat earth.

plicated—so much so that it is desirable to introduce the subject on the simplest possible basis. Accordingly, we shall assume for the moment that the propagation occurs over a flat earth. The incident radiation, during reflection from the earth's surface, suffers a phase change of ϕ radians and a reduction in amplitude to ρ times the incident value. We assume further that the direct and incident rays leave the radiator with the same amplitude, that is, the radiator has the same gain over the region containing the direct and reflected rays. The situation is illustrated in Fig. 175. The radiator and target are located at heights h_r and h_t m, and the distance between them is d m. The direct ray arrives at the target with an electric field strength E_0 volts per m. The reflected ray at the target has a field strength ρE_0 and is delayed in phase by ϕ radians plus β, the difference in electrical radians between the lengths of the direct and reflected rays. The total

phase delay is $\alpha = \phi + \beta$. The sum of the direct and reflected field strengths, at the target, is then

$$E_t = E_0 + \rho E_0 \epsilon^{-j\alpha} \tag{382}$$

where ρ is the reflection coefficient and the factor $\epsilon^{-j\alpha}$ is the imaginary exponential indication for a negative phase delay of α radians. From this expression we can find the pattern-propagation factor

$$F = \frac{E_t}{E_0} = |1 + \rho\epsilon^{-j\alpha}|$$
$$= \sqrt{1 + \rho^2 + 2\rho \cos \alpha} \tag{383}$$

Thus we have a relation for determining F from the reflection coefficient ρ and the total phase delay α. If the surface is smooth and highly conducting (such as sea water) and horizontal polarization is used, ρ is close to unity and ϕ is 180 deg. Then Eq. (383) becomes

$$F = \sqrt{2 + 2 \cos (180 + \beta)} \tag{384}$$

In this simple case we note that the value of F varies from zero (when β equals 0 or any even number of half wavelengths) to two (when β equals any odd number of half wavelengths). This is a most important conclusion, even though it has been derived for an idealized case: the maximum range of a radar may be reduced to zero by the effect of reflections or it may be increased to twice the free-space maximum range, depending on the nature of the reflection and the difference in length between the direct and reflected rays.

The most pronounced effect is that caused by the path difference. The path difference depends on the separation and relative heights of the radiator and target. Hence we should suspect that as the target approaches the radiator and the path difference changes, the echo signal strength at the target must change periodically from zero (or some small value) to a value well above the free-space value, the average of the periodic variations being the free-space value. The same effect is to be expected when the target height increases relative to the radiator height, or when the radiator height is changed and the target remains fixed. In other words, when ground reflections are present the vertical coverage diagram of a radar is not a single

lobe, as in the case of a high-gain antenna in free space, but a multilobed pattern, the number and extent of the separate lobes depending on the reflection conditions and the disposition of the radiator with respect to the earth's surface.

To show the nature of the vertical coverage-diagram in a simple case, we continue with the flat earth representation (Fig. 175), and compute the value of the path difference β from the geometry. On the assumption that d is large compared with h_r and h_t, it can be shown that the difference D in distance units between direct and reflected rays is approximately

$$D = (r_r + r_t) - r_d = \frac{2h_r h_t}{d} \tag{385}$$

The corresponding path difference in electrical radians is

$$\beta = \frac{4\pi h_r h_t}{\lambda d} \tag{386}$$

Then Eq. (384) becomes

$$F = \sqrt{2 + 2 \cos\left(\pi + \frac{4\pi h_r h_t}{\lambda d}\right)} \tag{387}$$

Since $\sqrt{2 + 2 \cos \alpha} = 2 \cos \alpha/2$, we can write this as

$$F = 2 \cos\left(\frac{\pi}{2} + \frac{2\pi h_r h_t}{\lambda d}\right) = 2 \sin \frac{2\pi h_r h_t}{\lambda d} \tag{388}$$

This formulation shows that as h_r, h_t, or d is varied, the other two remaining fixed, the pattern-propagation factor varies sinusoidally between the limits zero and two. A typical example of the vertical coverage diagram, plotted in conformance with Eq. (388), is shown in Fig. 176. The radiator height h_r is fixed and the values of h_t and d corresponding to points on the vertical coverage diagram inserted to obtain the lobe contours.

In the case of reflection from a poorly conducting surface, or one of rough outline, the foregoing assumptions regarding ρ and ϕ do not apply, and computations must be referred back to the basic equations, Eqs. (383) and (384). Since ρ is always a fraction, the excursions of F are not as great as in the case of the perfect reflector; that is, the signal is not reduced to zero in any direction, nor is it doubled in any other. But the general behavior is a quasi-sinusoidal variation, as suggested by Eq. (388).

We can now introduce the effect of the radiation pattern, which has been assumed isotropic in the preceding discussion. As shown in Fig. 177, we take the general case in which the

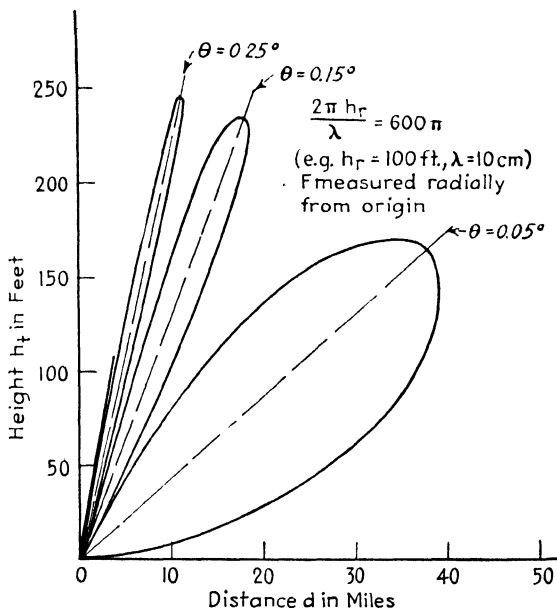



FIG. 176.—Vertical coverage diagram plotted from Eq. (388), showing lobes introduced by ground reflections. Note exaggeration of height scale.

direct ray as well as the reflected ray originate at angles off the main axis of the beam. We confine our attention to the vertical angle θ, and take θ_1 as the angle of the direct ray and θ_2 as that of the reflected ray. As shown, the direct ray has an amplitude

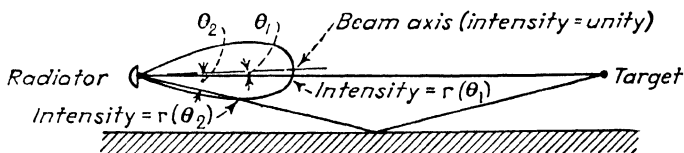



FIG. 177.—Effect of directive beam on relative intensity of direct and reflected components.

less than that along the main axis by the ratio $r(\theta_1)$ and the reflected ray is less by the ratio $r(\theta_2)$. Then the pattern-propagation factor is

$$F = r(\theta_1) + r(\theta_2)\rho\epsilon^{-j\alpha} \tag{389}$$

The ratio $r(\theta)$ is found from the vertical field strength directivity diagram of the radiator in question, by the methods of Eqs. (342) through (346). The symbol $F(\phi)$, used in those equations to express the polar diagram of directive arrays, corresponds exactly to the symbol $r(\theta)$ used in this discussion. Note however that the symbol F in Eq. (388) immediately above is not the same as the symbol F in Eqs. (342) through (346).

One further general observation may be made from the flat-earth equations just derived. Consider the pattern-propagation factor over a highly conducting flat earth, given by Eq. (388). In many cases, especially those involving the limiting distances of radar detection, the distance d is large compared with the heights of radiator and target. Then the sine in Eq. (388) may be replaced by its argument, with the result

$$F = \frac{4\pi h_1 h_2}{\lambda d} \tag{390}$$

Consider now the over-all effect of distance on the field strength. The field strength in empty space decreases as the first power of the distance. When reflections are present, the free-space field is multiplied by the factor F, which, as shown above, introduces another inverse first power of the distance. Thus the over-all field strength at the target decreases with the square of the distance. The power density at the target, which depends on the square of the field strength, then must decrease with the fourth power of the distance. Finally, the power delivered back to the radar receiver, involving two trips, must decrease with the eighth power of the distance. Thus when reflections are present, if the reflections are expressible by Eq. (390), the maximum range of a radar depends not on the fourth root of the transmitter power, receiver sensitivity, etc., but on the eighth root of these quantities.

To show this result in equation form, we return to Eq. (378), which gives the maximum range of a radar along a specified direction ϕ,θ, and substitute Eq. (379) in it, which gives

$$r_{max}(\phi,\theta) = \sqrt[4]{\frac{P_t G_0^2 F^4 \sigma \lambda^2}{P_{r\,min}(4\pi)^3}} \tag{391}$$

At the maximum range, F is given by Eq. (390) with $d = r_{max}$.

Substituting this in Eq. (391), we obtain

$$r_{\max}(\phi,\theta) = \sqrt[4]{\frac{P_t G_0{}^2 \sigma (h_1 h_2)^4 (4\pi)}{P_{r\min} r_{\max}{}^4 \lambda^2}} \tag{392}$$

and solving for $r_{\max}$

$$r_{\max} = \sqrt[8]{\frac{P_t G_0{}^2 \sigma (h_1 h_2)^4 (4\pi)}{P_{r\min} \lambda^2}} \tag{393}$$

This "eighth-root" law, it must be remembered, does not have the general application of the free-space radar equation (fourth-root law), since it has been derived only for propagation over a highly conducting flat earth at such large distances or such small heights that the sine of Eq. (388) may be replaced by its argument. The eighth-root law applies to cases where h_1 and h_2 are small and d and λ are large, provided that d is not so large as to invalidate the flat-earth assumption. In short, no such simple formulation as Eq. (393) suffices in any general sense to give the maximum range of a radar when reflections are present. Rather the complete formulation of the pattern-propagation factor is required, and a vertical coverage diagram must be constructed for the particular wavelength and radiator height in question. Moreover, the curvature of the earth cannot be neglected except over very short ranges. For a more complete treatment we must consider the geometry of propagation over a spherical earth.

103. Propagation over Reflecting Spherical Earth.—When we consider the spherical earth, we are confronted by the horizon, a concept not encountered in the flat-earth analysis. The radio horizon is the boundary defined by the points of tangency between the radio lines of sight and the earth's surface. Note that "radio" line of sight, not "optical" line of sight, is meant. The radio line of sight is generally greater than the optical, as we shall show later in this section.

The basic concept is the optical line of sight illustrated in Fig. 178. As the figure shows, the radius of the earth R plus the radiator height h_r form the hypotenuse of a right triangle, of which the radius and the line of sight are the other sides. Since in all cases the height of the radiator is small compared with the earth's radius, we can neglect its square and obtain

$$r^2 = 2Rh \tag{394}$$

from which

$$r = 1.23 \sqrt{h} \quad \text{miles} \tag{395}$$

when h is expressed in feet. This expression gives the optical horizon distance from any elevated object. In the case of an elevated radar radiator viewing an elevated target, the maximum unobstructed distance d_{max} between them occurs when both view a common point of tangency (Fig. 178), and

$$d_{max} = 1.23 (\sqrt{h_t} + \sqrt{h_r}) \quad \text{miles} \tag{396}$$

where h_r and h_t are the heights in feet of the radiator and target.

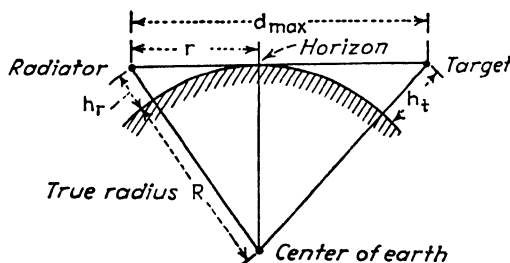



FIG. 178.—Geometry of optical horizon and maximum unobstructed distance between elevated points.

These purely geometrical lines of sight must be adjusted to take account of refraction and diffraction when they are applied to radio wave propagation. Refraction varies between wide limits with changes in the condition of the atmosphere and diffraction with the nature of the surface near the tangent point, and thus no hard and fast rules can be laid down. One generalization is of some value. The change of index of refraction with height in the normal atmosphere in the temperate zones produces a bending of the propagated ray whose radius of curvature is about four times that of the earth's surface. As shown in Sec. 105, if we assume a fictitious radius $R' = \frac{4}{3}R$ and plot the ray as a straight line, we obtain the actual point of tangency. Then Eq. (394) becomes

$$r^2 = \tfrac{8}{3}Rh \tag{397}$$

and

$$d_{max} = 1.41(\sqrt{h_t} + \sqrt{h_r}) \quad \text{miles} \tag{398}$$

This increase in the line of sight is ordinarily to be expected due to atmospheric refraction, but this expectation may go completely unrewarded in exceptional weather.

Beyond the radio line of sight, propagation is possible by diffraction, but the attenuation of the signal beyond the radio horizon is generally very rapid, when only diffraction is present. Exceptional propagation beyond the horizon is explained in terms of guided waves, trapped between a low-lying refracting layer and the surface. An account of such anomalous propagation is given in Secs. 104 to 106.

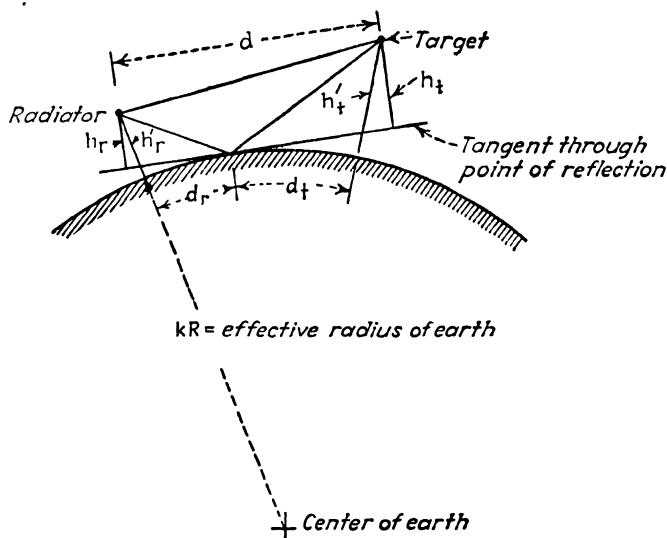



FIG. 179.—Geometry of radio-wave propagation over spherical earth. The flat-earth representation (Fig. 175) is shown tangent at the point of reflection.

The vertical coverage diagram of a radar over a spherical earth, within the horizon distance, can be computed by reference to the flat-earth representation previously discussed. As shown in Fig. 179, the earth is represented as a sphere of radius kR, where R is the true radius (= 6,400 km) and k is the factor (usually taken as $4/3$) which accounts for the usual refractive bending. The spherical geometry is converted to plane by passing a tangent through the point of reflection and erecting the perpendicular heights h_r and h_t to radiator and target as shown. These heights are related to the true heights h_r' and h_t' by the approximate expression

$$h_r = h_r' - \frac{d_r^2}{2kR} \tag{399}$$

for the radiator and by a similar expression for the target.

One further conversion is required, that is, from the horizontal direction at the radiator to the direction of the tangent. The difference between these two directions is the angle θ_d, which is given by

$$\theta_d = \frac{d_r}{kR} \quad \text{radians} \tag{400}$$

Using these substitutions, the vertical coverage diagram may be

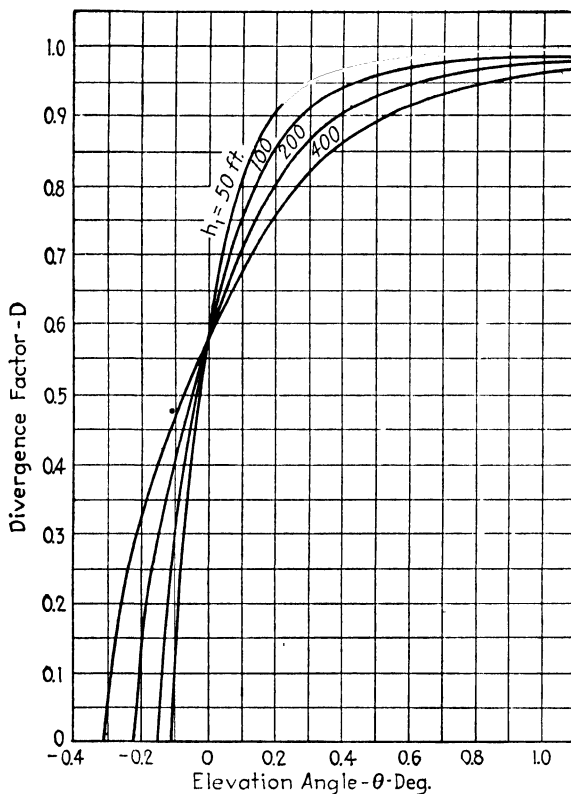



FIG. 180.—Divergence factor for Eqs. (403) and (404), for various transmitter heights. (*After Frank and Chu.*)

constructed by reference to the flat-earth theory of the previous section.

The pattern-propagation factor may also be computed directly by the use of the following expression [cf. Eq. (383)]

$$F = \sqrt{1 + (D\rho)^2 + 2D\rho \cos \alpha} \tag{401}$$

where D is a divergence factor that modifies the reflection coefficient ρ to take account of the spherical surface. D is given by the approximate expression

$$D = \frac{1}{\sqrt{1 + (2d_r{}^2 d_t / kRh_r{}'d)}} \tag{402}$$

where the symbols are given in Fig. 179.

Evidently the computation of vertical coverage diagrams is not a simple matter when carried out in detail. For many purposes it is sufficient to construct an approximate coverage diagram in the form of the *envelopes* of the maxima and minima of the lobes in the coverage diagram. The maximum envelope is given by

$$r_{max} = r_0(1 + D\rho) \tag{403}$$

and the minimum envelope by

$$r_{min} = r_0(1 - D\rho) \tag{404}$$

where r_0 is the free-space maximum range. These envelopes may be constructed when the maximum range is 100 miles or greater in terms of the vertical angle and the radiator height only. Figure 180, computed by Frank and Chu, gives values of D for angles θ from -0.4 to $+1.0$ deg and radiator heights from 50 to 400 ft above the level of the reflecting surfaces. It is clear from the equations that the free-space range r_0 is the arithmetic mean, $(r_{max} + r_{min})/2$, of the maximum and minimum envelopes.

The implications of the discrete lobes within the vertical coverage diagram are fairly evident. A radar may fail to detect a target if the target lies within a minimum, whereas it will detect at great ranges if the target lies within a maximum. To smoothen this irregular performance, it is desirable to fill in the coverage diagram with as many lobes as possible, and to avoid ground reflections as much as possible. Both effects may be achieved by the use of h-f radiation. The use of short wavelengths (high frequencies), as reference to Eq. (388) shows, produces a large number of closely spaced lobes, and thus the unfilled gaps are not extensive. Moreover, by the use of short wavelengths it is possible to produce narrow beams that may be directed close to the horizon, in search of low targets,

without delivering sufficient energy to the earth's surface to cause deep minima in the coverage diagram.

The most important effect of ground and sea reflections is the limitation on "low angle" coverage, that is, the short detection range obtainable in the horizontal direction and below that direction. This effect is caused by the universally present pronounced minimum in the lobe pattern at zero elevation, shown in Figs. 176 and 181 and evident in Eq. (388).

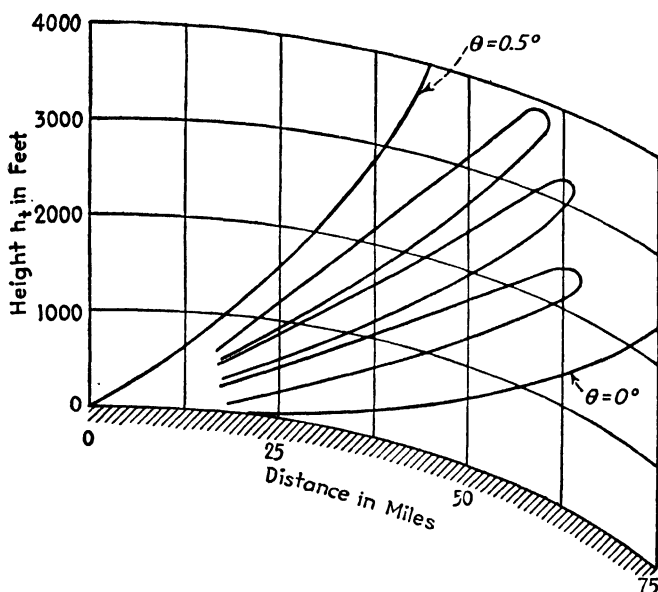



FIG. 181.—Typical vertical coverage diagram plotted on curved coordinates to reduce distortion.

104. Anomalous Propagation of Radar Signals.—The foregoing treatment of radar propagation over a spherical earth has been carried out on the assumption of passage through the "standard troposphere," which introduces bending of the rays equivalent to a one-third increase in the radius of the earth. Experience has indicated that departures from the standard troposphere exist frequently, especially in tropical regions, and that such departures may have a pronounced effect on the maximum detection range of a radar.

The more spectacular occurrences are those of long-distance detection, such as the detection of a coastline across the Arabian

Sea at a distance of 1,750 miles by a standard 200-megacycle radar, which was reported in 1944. The reverse effect also occurs, although less frequently, when the normal detection range is reduced by a factor of 2 or more for no reason connected with the radar itself or the target.

An important practical effect of such anomalous propagation is the introduction of clutter areas on the radar indicator which may obscure targets which would otherwise be visible. If, for example, the normal maximum range of a radar is 100 miles and an indicator of 100 miles maximum range is in use, clutter from beyond 100 miles, picked up under anomalous conditions, will appear on the next subsequent sweep and will fall directly over the nearer targets. Long-distance detection of targets is equally possible, of course, under the same conditions; but since the anomalous occurrences are difficult (though not impossible) to predict and are impossible to control, operational procedures are rarely set up to take advantage of exceptional atmospheric conditions.

The exceptional ranges caused by anomalous propagation are confined to paths within 1 deg of the horizontal, as viewed from the radiator. Hence radars which search at high angles and airborne radars which search well above or below the horizontal are not affected, whereas ground-based or shipborne radars used to search horizontally for shipping or low-flying aircraft frequently display abnormal maximum ranges. Anomalous propagation has received concentrated attention in an attempt to devise means of predicting its occurrence. The results of this study are reviewed briefly in the following sections.

105. Refractive Bending of Wave Paths.—We have previously mentioned that refractive bending of radio waves in the standard atmosphere may be taken into account by using an effective radius of the earth equal to $4/3$ times its actual value and by plotting the wave path as a straight line. We shall now go into this concept more in detail and extend it to the case of anomalous propagation.

Refractive bending of wave paths results from the decrease in the refractive index of the atmosphere with height. The physical basis of this effect is readily seen from the fact that the refractive index depends on the dielectric constant of the atmosphere, which in turn depends upon the density of the atmospheric

components, oxygen, nitrogen, water vapor, etc. In the normal atmosphere, these components become less dense as the height increases.

The refractive index of the atmosphere is defined as the ratio of the velocity of radio propagation in vacuum to the velocity in the atmosphere. At sea level, the value of the index is approximately 1.0003, and it decreases with height at a rate of about -1.2×10^{-8} per ft. The waves thus travel faster in the upper atmosphere than they do near the surface. Hence the path of any wave that starts horizontally is bent downward towards the earth. When such bending is present, certain of the bent waves may reach the earth beyond the optical horizon,

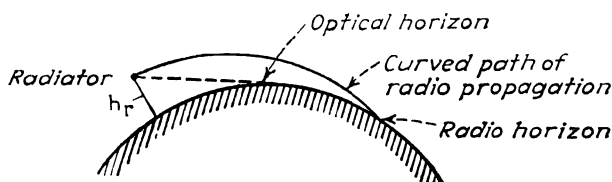



FIG. 182.—Extension of the radio horizon beyond the optical horizon, caused by refractive bending of the radio path.

as illustrated in Fig. 182. When the degree of bending is sufficiently pronounced and localized, waves leaving at a certain angle are bent parallel to the curvature of the earth and may continue parallel to the earth for a distance many times the optical horizon distance. This phenomenon is known as "trapping" and the region within which the waves travel is known as a "duct." The waves are propagated within the duct in much the same manner as in a waveguide.

We shall consider first the normal case in which the index of refraction n decreases linearly with height h at a rate $dn/dh = -1.2 \times 10^{-8}$ per ft. The radius of curvature of the path of a wave passing through such a gradient of index is given by

$$\rho = -\frac{1}{dn/dh} \tag{405}$$

or about 8×10^7 ft. This is about four times the actual radius of the earth (2.1×10^7 ft). If we subtract the curvature $1/\rho$ of the wave path from the curvature of the earth's surface $1/R$, we obtain the curvature of a fictitious earth $1/kR$ over which the waves may be represented as traveling in straight lines. So

long as the index gradient dn/dh is a constant, this simplified picture holds. To obtain the value of k in terms of the gradient and the radius we equate the curvatures

$$\frac{1}{R} - \frac{1}{\rho} = \frac{1}{kR} \tag{406}$$

substitute Eq. (405) and solve for k

$$k = \frac{1}{1 + R\, dn/dh} \tag{407}$$

When dn/dh has the standard value -1.2×10^{-8} per ft, and and $R = 2.1 \times 10^7$ ft; then $k = 1/(1 - 0.25) = 4/3$, which is the equivalent radius factor used in the preceding section.

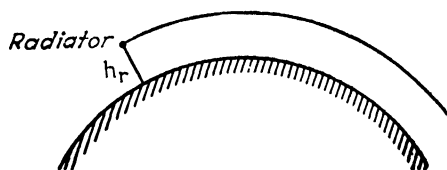



FIG. 183.—Curvature of refractive bending equal to that of earth's surface. The radio horizon is then nonexistent.

Under certain abnormal conditions, known as "substandard," the index will fall off with height more rapidly than the standard value. Then the value of k increases in accordance with Eq. (407). For example, if $dn/dh = -3.6 \times 10^{-8}$ per ft, $k = 4$, and if $dn/dh = -4.8 \times 10^{-8}$ per ft, $k = \infty$. This latter case is of particular interest since the effective surface of the earth then becomes flat, and all waves directed horizontally from an elevated radiator are not intercepted by the earth. Such a gradient produces a bending of horizontally directed rays just parallel to the earth's surface (Fig. 183).

In practice the gradient of the index is not constant, but varies nonlinearly with height. In such cases the effective radius, given by Eq. (407), is no longer constant and the straight-line representation of the paths loses its convenience. It is then convenient to define a new index N known as the "modified relative index of refraction," such that

$$N = \frac{n}{n_0}\left(1 + \frac{h}{h_0 + R}\right) \cong \frac{n}{n_0} + \frac{h}{R} \tag{408}$$

where n is the index at height h, n_0 is the index at the height h_0 (the height of the radiator), and R is the radius of the earth. It should be noted that N normally increases with height because the term h/R overcomes the decrease of n with height, previously discussed. Moreover, the rate of change of N with height, dN/dh, is approximately the curvature of the effective earth $1/kR$, that is,

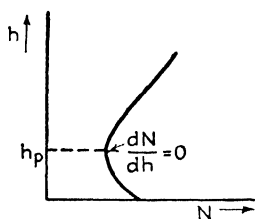



FIG. 184.—Typical plot of modified index of refraction against height above the earth.

$$\frac{dN}{dh} = \frac{1}{kR} \tag{409}$$

If we plot the curve of N against height and find, as in Fig. 184, that the slope dN/dh of the curve at a given height h_p is zero, the effective radius at that point is infinite. Any horizontally directed waves at that height will be bent so as to follow the curvature of the earth. Thus the height at which dN/dh is zero is the height of the top of the duct, and a duct can exist only so long as dN/dh is zero somewhere along the curve.

If we wish to trace the path of a wave, point by point, in a given situation, we may do so conveniently in the flat-earth representation shown in Fig. 185. The effective earth repre-

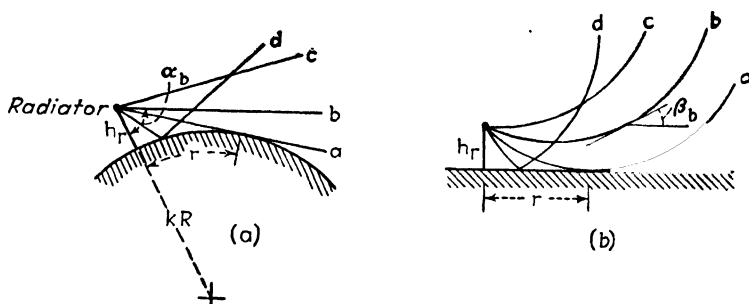



FIG. 185.—Transformation from spherical-earth (a) to flat-earth (b) representation of ray paths.

sentation, shown in Fig. 185a, is constructed on a sphere of radius kR, with straight-line wave paths leaving the radiator at various angles α relative to the vertical. The flat-earth transformation (Fig. 185b) is obtained by subtracting the effective curvature $1/(kR)$ of the earth from the surface (thereby obtaining a flat surface) and from each path (thereby causing

the paths, rays to appear concave upward with radii kR). The angle β between the horizontal and the curved ray path, at any height, is given by

$$\beta = 1.41 \sqrt{N - \cos \alpha} \qquad (410)$$

where N is the value of the modified index at the height in question. To determine whether a given wave will be bent so as to propagate beyond the horizon, it is possible to trace the wave path by this method and note the distance at which it strikes the earth, relative to the horizon distance computed by Eq. (398).

106. Computation and Measurement of Modified Index Curves.—It is clear that the tracing of wave paths, as well as the simpler matter of determining whether or not a duct exists, depend on the modified index plotted against height. The plotting of such curves has become something of a science, and a standardized procedure has been developed for the purpose. In the first place, N is not used as a variable in the curves, because it is inconvenient, consisting of 1 plus a very small fraction (1.000001 at sea level under normal conditions). In its place another quantity, known as $M - M_0$ is defined as

$$M - M_0 = (N - 1) \times 10^6. \qquad (411)$$

The value of $M - M_0$ at sea level is about unity. The quantity M_0 is related to the index of refraction n by

$$M_0 = (n - 1)10^6 \qquad (412)$$

Its value is about 300 at sea level. The value of M increases from about 300 (plus 1) at sea level to about 400 at several thousand feet. Differentiation of Eq. (411) shows that dN/dh is proportional to dM/dh; therefore a curve plotted in terms of $M - M_0$ has the same shape as one plotted in terms of N. In particular, the values of zero slope occur at the same height in the two cases.

With these definitions in mind, we turn now to the computation and measurement of the M curves. The computation relates fundamentally to the values of the basic index n at various heights. An empirical formula gives the value of n at a particular

height when the pressure temperature and vapor content of the atmosphere are known at that height.

$$n = 1 + \frac{79}{T}\left(P - e + \frac{4{,}800e}{T}\right) \times 10^{-6} \qquad (413)$$

where P is the total pressure in millibars, T the temperature in degrees Kelvin, and e the vapor pressure in millibars, all at the height in question. These quantities may be determined

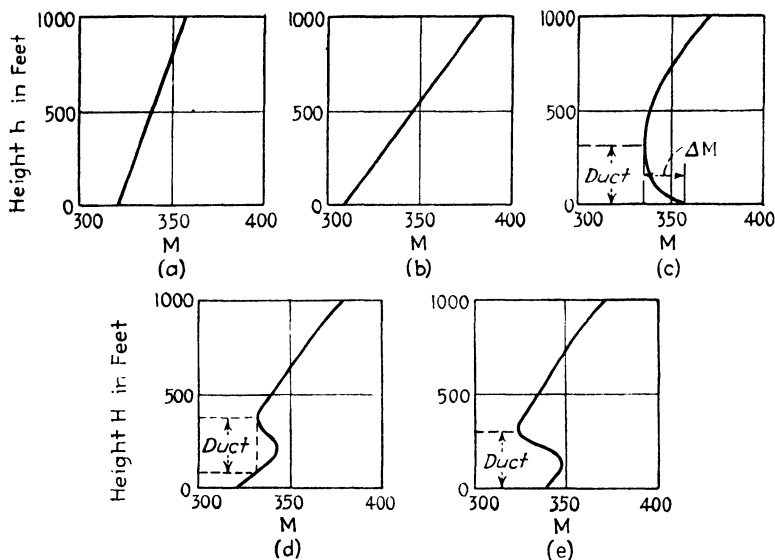



FIG. 186.—Typical modified index curves: (*a*) standard atmosphere, (*b*) substandard, (*c*) C-shaped, (*d*) S-shaped, (*e*) ground-based S-shaped.

as functions of height by sounding the atmosphere with a radiosonde (temperature and pressure instruments borne aloft by a balloon and fitted with a radio transmitter for reporting). The values at the surface may be found directly, thus establishing n_0 and M_0 by Eqs. (413) and (412), respectively. From n as a function of height, and from Eq. (408), N is found as a function of height, and from N, by Eq. (411), $M - M_0$ is found as a function of height. This last function is plotted and used as the basis of the ensuing analysis. Typical M curves are shown, in somewhat idealized form, in Fig. 186. Figure 186*a* is the normal M curve of the standard atmosphere, in which dM/dh has a positive value of about 3.6 per 100 ft. As we have

seen, the bending associated with this standard gradient has a radius of curvature about four times the earth's radius, and may be represented as a straight line over an effective earth of radius $4R/3$. Figure 186*b* is a substandard curve, in which dM/dh has a greater, but still positive, value. This indicates a less rapid decrease of n with height and less intense bending. The corresponding value of k is decreased, and the radio horizon is closer than under normal conditions. There is, therefore, no trapping of radiation. Figure 186*c* is the so-called "C-shaped curve," which has a negative slope below a certain height and a positive slope above it. Where the slope goes through zero the conditions for trapping are satisfied, as previously discussed (page 274), and a duct exists from this height downward to the surface. This condition is known as "surface-trapping." Figure 186*d* is the so-called "S-shaped curve," with two values of zero slope. The upper point of zero slope indicates the height of the top of the duct and the point directly beneath it indicates the height of the lower limit of the duct. This is an elevated duct. An intermediate case, Fig. 186*e*, is the ground-based S curve, in which the duct extends to the earth's surface.

In addition to giving the heights at which trapping occurs, these curves indicate indirectly the limiting angle α, above the horizontal, at which the wave may be directed from the radiator and yet be subject to trapping. From the value of $M - M_0$ at the point of zero slope we compute by Eq. (411) the value of N at that height and substitute this value with $\beta = 0$ in Eq. (410). This gives the limiting value of angle α_p for trapping. Waves directed above this angle are not trapped, and those below are trapped and returned to the earth. Rays emitted at the angle α_p travel parallel to the earth's surface indefinitely so long as the trapping condition persists, until the dispersion of the signal laterally weakens it beyond recognition. Figure 187 gives the limiting angle for various values of ΔM. It will be noticed that the limiting values are under 40 min of arc.

The foregoing analysis has been carried out without reference to the wavelength of the propagated radiation. It can be shown that radiation is trapped and propagated with small loss only when the height (vertical dimension) of the duct is several hundred times as great as the wavelength. Thus a duct height

of about 1,000 ft is required to trap and transmit appreciable amounts of energy at 1-m wavelength, whereas duct heights of 100 ft suffice for 10-cm wavelength. Since the smaller ducts are more commonly encountered, centimeter waves (S, X, and K bands) are trapped more frequently than meter waves (P and L bands). The maximum wavelength that can be trapped efficiently is given by

$$\lambda_{max} = 2 \times 10^{-3} h_d \sqrt{\Delta M} \quad \text{cm} \qquad (414)$$

where h_d is the height of the duct in feet and ΔM is the displacement of the point of zero slope from the base of the curve.

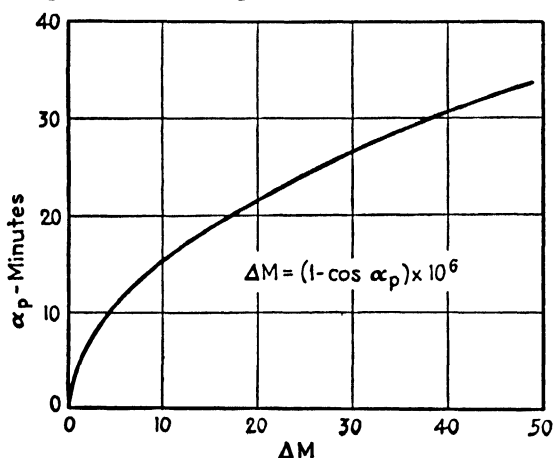



FIG. 187.—Limiting angles for trapping.

107. Relation of M Curves to Weather Phenomena.—The detailed procedure used in predicting anomalous propagation from meteorological data is beyond the scope of the present treatment, but a few qualitative relations may be of interest. Conditions favorable to the formation of a duct occur usually in fine weather, when the upper atmosphere (up to 1,000 ft) is abnormally dry and warm. These conditions reduce the density, and hence the refractive index, of the upper strata. On the other hand if the upper atmosphere is exceptionally cold, the density and refractive index are above normal and no anomalous conditions are to be expected. Stormy, turbulent weather usually produces a mixed condition of the atmosphere that destroys the duct even if conditions are otherwise favorable to its formation. Turbulence is more apt to occur over land

than over water; and as a result anomalous effects are more frequently observed on over-water paths, and this is particularly true during hours of sunlight. After darkness, the surface inland cools rapidly by radiation, whereas the atmosphere several hundred or thousand feet up retains its warmth. This is a condition favorable to the formation of a duct. Exceptional range of detection over land is thus apt to occur most frequently at night, in clear weather, and particularly when the surface temperature drops sharply with the coming of nightfall. Water areas to the leeward of land masses often display anomalous conditions, due to the drifting of warm dry air from the land over the cooler moist air near the water. The prevalence of clear weather in certain tropical regions accounts for the prevalence of anomalous propagation during dry seasons. On the other hand, rainy seasons in the equatorial zone are unfavorable.

108. Absorption of Signals by the Atmosphere[1].—Our treatment of radio propagation up to this point has assumed attenuation of the signal due to one cause only, the dispersion of the energy as it passes through space. In free space the dispersion of the energy is represented on the surface of an expanding sphere. When ground reflections are present, the same law of dispersion is assumed to affect the direct and reflected waves.

A second cause of attenuation, absorption and scattering by the atmosphere, occurs under certain conditions. The most generally encountered atmospheric attenuation is caused by rainfall in the path of the wave. Attenuation is also caused, at very short wavelengths, by resonances within the molecules of oxygen and water vapor.

In computing the attenuation we shall neglect dispersion and assume that the power P leaving one face of a unit cube is equal to the power P_0 entering the opposite face, less the loss due to absorption and scattering within the volume. Then the two powers are relating by an exponential law with distance

$$P = P_0 \epsilon^{-\alpha s} \tag{415}$$

[1] MUELLER, G. E., Propagation of 6-cm waves, *Proc. I.R.E.*, **34,** 181P (April, 1946).

BERINGER, R., Absorption of ½-cm electromagnetic waves in oxygen, *Phys. Rev.*, **70,** 53 (July 1, 1946).

ROBERTSON and KING, Effect of Rain on Propagation of Waves in the 1-cm and 3-cm Regions, *Proc. I.R.E.*, **34,** 178P (April, 1946).

where α is the attenuation constant and s is the distance through which the energy passes in traversing the cube. It is convenient to express the attenuation in decibels.

$$db = 10 \log_{10} \frac{P}{P_0} = -4.34\alpha s \tag{416}$$

This expression shows that the attenuation, expressed in decibels, is proportional to the distance traveled and is equal to 4.34α db per unit distance.

The value of the attenuation constant depends, evidently, on the density of the absorbing medium, as well as on many other factors relating to the composition of the drops or molecules, such as their size and separation, internal resonance frequencies, etc.

Attenuation due to rainfall has been measured by various workers. Their work indicates that the attenuation is approximately proportional to the density of the rain, which can be measured in terms of the precipitation rate in millimeters per hour. The attenuation constant also depends on the size and separation of the droplets but the variations due to these effects are difficult to identify. Over the range from a light drizzle (1 mm per hr) to a cloudburst (100 mm per hr), the attenuation in decibels per mile increases linearly.

TABLE VII.—ATTENUATION CONSTANTS FOR RAINFALL

Band	Frequency, mc	Wavelength, cm	Attenuation, db/mile per mm/hr
S	3,000	10	0.004
X	10,000	3	0.05
K	30,000	1	0.3

The attenuation constant changes also with the frequency of radiation. Over the range from 1-cm to 10-cm wavelength (the range within which rainfall attenuation is of practical importance), the attenuation constant varies approximately as the square of the frequency. Table VII gives approximate values of attenuation constants in decibels per nautical mile per millimeter per hour for S, X, and K bands.

It should be realized that the measured values of these constants display wide variations, due to the experimental difficulties and uncertainties concerning the distribution and constitution of the rainfall over the propagation path.

The effect of rainfall attenuation is greatest at the high frequencies, when the precipitation is heavy and when the distance from the target to the radar is great. The latter effect is caused by the long path through which the attenuation occurs. To examine this concept quantitatively, consider a radar viewing a given target in clear weather at the maximum range, r_c miles, at which the given target is visible. Consider next the maximum range of the radar r_r miles on the same target when rain falls at a rate p mm per hr. In the latter case two effects are present: (1) The rainfall causes attenuation equal to $2apr_r$ db where a is the attenuation constant in db/mile per mm/hr, p is the precipitation rate in mm per hr, and $2r_r$ is the distance in miles travelled under the rainfall condition. (2) The reduction of range from r_c miles in clear weather to r_r miles in rainy weather introduces a potential increase in the reflected power received by the radar by a factor $(r_c/r_r)^4$, or a power increase of $40 \log_{10} r_c/r_r$ db, assuming that the fourth power law is operative (free-space propagation). This potential increase in received power is just balanced by the attenuation due to rainfall, at the range r_r. Hence we have the equality

$$2apr_r = 40 \log_{10} \frac{r_c}{r_r} \qquad (417)$$

The maximum range under the given condition of rainfall is then

$$r_r = \frac{20 \log_{10} r_c/r_r}{ap} \qquad (418)$$

A plot of this equation giving the maximum range to be expected under given conditions of rainfall and frequency is shown in Fig. 188. It will be noted that the decrease in maximum range when rain is present is much more pronounced, relatively speaking, when the maximum range in clear weather is great.

It should be remembered that this analysis deals only with the attenuation introduced by the rainfall. An additional, and even more pronounced, effect is the back scattering of energy toward the radar, from the rain itself. A rainstorm becomes a

potent source of clutter when such back scattering is present, and the clutter (much like random noise in appearance on the indicator screen) makes difficult the detection of targets, even when the target is well within the maximum range r_r predicted by Eq. (418).

Absorption due to the molecular components of the atmosphere, including water vapor, is generally not prominent except at the very shortest wavelengths (0.5 and 1.0 cm wavelength).

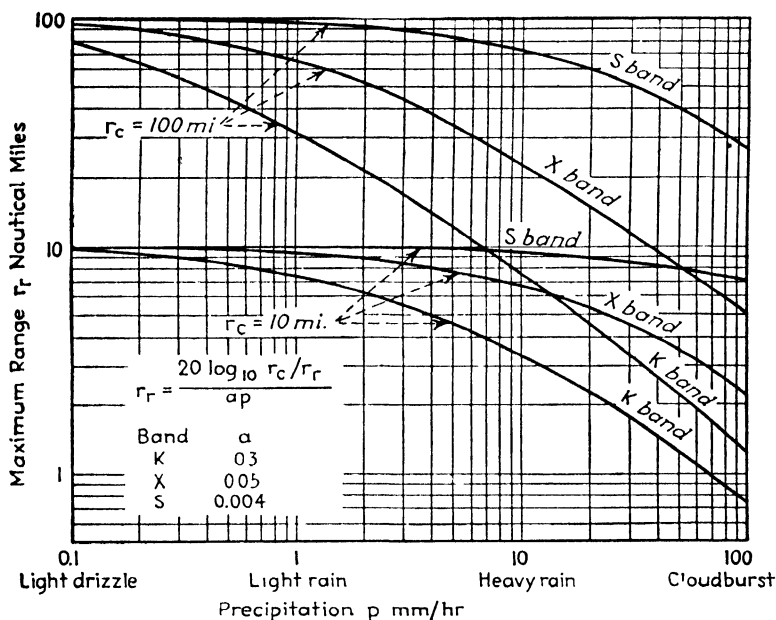



FIG. 188.—Effect of precipitation on maximum range of radar detection.

In the K band, the molecular absorption is of the order of 0.1 db per mile. At the K/2 band, at about 0.5 cm, however, a prominent resonance band occurs in the oxygen spectrum, which introduces a maximum attenuation of the order of 10 db per mile. This severe attenuation limits the utility of wavelengths in this vicinity, for radar applications, to detection at not more than a few miles.

109. The Radar Target.—Before the performance of a radar system may be specified, account must be taken of the properties of the target. In the preceding discussion we have characterized the target by a number, its echo area (radar cross section). In

this and the following sections we examine the concept of the echo area in further detail and discuss other aspects peculiar to the target, such as the effect of immediate surroundings and motion relative to the radar.

As we recall from Chap. I (page 41), the echo area of a target is a fictitious area that replaces the actual target. The echo area is imagined as intercepting an amount of power which, if reradiated equally in all directions, would produce at the radar receiver an echo equal to that actually observed from the given target. Suppose the power density that falls on the target is S_t and the echo area of the target is σ. Then the power intercepted is $S_t\sigma$. If this power were reradiated uniformly in all directions, it would produce a power density at the receiver of $S_t\sigma/4\pi r^2$ where r is the distance between radar and target. Thus if S_r is the power density actually observed at the receiver, the echo area may be computed from

$$\sigma(\phi,\theta) = 4\pi r^2 \frac{S_r}{S_t} \tag{419}$$

To compute σ empirically, it is necessary to measure or compute the incident power density S_t, the distance r, and the received power density at the receiver S_r.

From a physical point of view it is evident that the echo area depends on the aspect of the target presented to the radar. Thus a dipole, considered as a target, has zero area when viewed end-on, because a dipole radiates no power in the direction of its length. The echo area of a dipole is a maximum, on the other hand, when viewed broadside. The dependence of σ on the target aspect is indicated in Eq. (419) by the variables ϕ and θ. These represent the direction of the line connecting radar and target with respect to a reference direction in the target (usually taken as the direction of maximum echo area). In general the echo area of a target may be represented in three dimensions by a figure, similar to the directivity diagram of a radiator.

A target whose echo area may be simply computed is the uniform conducting sphere. When the diameter of the sphere is large compared with a wavelength, the distribution of the reradiated energy is very nearly uniform in all directions (isotropic). When the power density incident along the target-radar direction is S_t, the power intercepted is S_t times the maximum

cross section of the sphere πa^2, where a is its radius. The power reradiated, assuming no power lost in the sphere, is then $\pi a^2 S_t$, and the power density at the receiver is $S_r = \pi a^2 S_t / 4\pi r^2$. Substitution in Eq. (419) gives

$$\sigma = \pi a^2 \tag{420}$$

The echo area of a large sphere is thus equal to its maximum cross section.

Another simple form is the dipole target. The radiation pattern of a dipole is the familiar figure-eight pattern (Fig. 150). The echo area increases rapidly as the length of the dipole approaches resonance (one-half wavelength). Dipole targets are occasionally encountered in practice, usually in the form of half-wave conductors ("chaff") scattered by aircraft to confuse enemy radar-warning systems. The half-wave conductor reradiates all the energy abstracted from the incident wave, and its directivity diagram is the square of the radiation pattern of a half-wave dipole. When viewed broadside, the echo area of a half-wave conductor is

$$\sigma = \frac{3\lambda^2}{8\pi} \tag{421}$$

If the half-wave conductor is cut in the center and connected to a resistance equal to its radiation resistance, the same value of echo area applies. The amount of energy absorbed increases in this case, but the additional energy is lost as heat in the resistor and is not reradiated.

As would be suspected from the discussion of dipole sheets, (Sec. 91) a flat metallic target, viewed perpendicularly, has a highly directional reradiation pattern, and the echo area along the normal direction increases very rapidly as the size of the sheet increases. The echo area along the normal is

$$\sigma = \frac{4\pi l^4}{\lambda^2} \tag{422}$$

where l is the length of the edge of the sheet and λ is the wavelength.

Another form of target that permits calculation is the so-called "corner reflector." This target, analogous to the corner reflector

used in optics, has the property of reradiating a maximum signal in the direction from which it was illuminated, irrespective of its aspect relative to the radar. The corner reflector is useful as a "marker" target on the ground, since it can be viewed by several airborne radars simultaneously and return a strong signal to each. A corner reflector has a maximum echo area along the axis of symmetry, that is, along the line which makes equal angles with each of the edges The echo area of a triangular corner reflector (composed of three faces, Fig 189) whose area is large compared with a square wavelength is

$$\sigma = \frac{4\pi l^4}{3\lambda^2} \qquad (423)$$

where l is the length of the edge Corner reflectors must be accurately and rigidly constructed if they are to preserve their directive property.

FIG 189 —Corner reflector

In most practical radar work, of course, these specialized geometrical forms are not encountered. In their place are found targets, such as aircraft and surface ships, that do not possess simple geometry. Theoretical computation is impossible in such cases, and it is necessary to resort to empirical studies Statistical analysis of measurements, with a target of fixed size and shape, will reveal the average echo area of the target at a particular frequency and the variations from the average as the target aspect changes The experimental difficulties in such a program are numerous, particularly if the target is an aircraft, since its aspect relative to the target-radar line is seldom known with precision Moreover the evaluation of the echo area of a target is complicated by surface reflections, and these complications approach the bizarre when the target has a many-lobed reradiation pattern As a result, knowledge of target echo areas, as measured directly from practical targets under controlled conditions, is very meager Table VIII gives average echo areas generally assigned in computing maximum free-space ranges of S-band radar equipments

TABLE VIII.—AVERAGE ECHO AREAS AT S BAND

Type and Size of Target	Echo Area, sq m
Fighter	10
Medium bomber	25
Heavy bomber	50
Very heavy bomber	150
Destroyer	500
Cruiser-freighter	1000
Battleship	2000
Submarine conning tower	5
81-mm mortar shell	1
Chaff (1,000 dipoles dispersed through one cubic meter)	50

110. Target Contrast. Effect of Target Surroundings.—When a radar target is viewed against a nonreflecting background (for example, an aircraft viewed against the sky), the visibility of the target is limited only by the maximum range of the radar, computed in accordance with the principles previously outlined. In many circumstances, however, the target is displayed against a reflecting background Examples are low-flying aircraft at or near the horizon, ground targets observed by airborne radars, and surface targets observed by ground-based or shipborne radars. When background reflections are present, the visibility of the target is usually limited, not by the power and sensitivity of the radar equipment, but by the difficulty of distinguishing the target from its background

Distinction of a target from its background is based on either of two fundamental properties: the target may have a higher (or lower) radar reflectance than the background, or it may be in motion relative to the background.

In this section we consider distinction of targets by difference in reflectance, that is, by target contrast. Target contrast is said to exist when the current induced in the target by the incident radar signal is substantially greater (or less) than that induced in the background, and when the aspect of the target is such that this difference is preserved in the reradiated signal. In military applications, for example, the target is ordinarily a metal structure. Its conductivity and hence the reradiated signal are generally considerably higher than those of its background.

When target contrast is available, the degree of distinction between target and background depends on the relative illumina-

tion received by each. To emphasize the contrast, it is necessary that the radar beam illuminate the target fully and the background as little as possible. The restriction of background illumination, as shown in Fig. 190, must be three-dimensional when the background surface is viewed obliquely, as is generally the case. Thus it is necessary to employ a narrow beam to restrict illumination in the elevation and azimuth coordinates and a short pulse to restrict illumination in the range coordinate. The optimum case occurs when the "volume" of the pulse (cross-sectional area of the beam at the target multiplied by the depth of the pulse in space) is substantially equal to the volume of the target. As the space occupied by the pulse becomes smaller than this, no improvement in contrast appears, but as it becomes larger the background illumination increases and may overshadow the echo from the target.

FIG. 190.—Effect of pulse width on illumination of area surrounding target.

When the target is small, or viewed at a great distance, it is virtually impossible to avoid background illumination. Thus the cross section of a beam 1 deg wide (narrow as radar beams go) is 460 ft in diameter at a distance of 5 miles, and the depth in space of a 0.25 μsec pulse (short as radar pulses go) is about 245 ft. These dimensions exceed those of all but the largest targets.

111. Targets in Motion. Doppler Effect.—Targets may be distinguished from reflecting backgrounds, even in the absence of radar contrast, when the target is in motion relative to the background and the radar. Detection in this case is based on the shift in frequency due to the Doppler effect that occurs during reflection of the signal from the moving target. The amount of the Doppler frequency shift is minute relative to the carrier frequency, since the speed of the target is always very small compared with the velocity of propagation of the signal. But, since the carrier frequencies employed in radar are high, appreciable shifts of frequency (in the a-f range) are produced by relative speeds of tens or hundreds of miles per hour.

To those unfamiliar with the Doppler principle, the following derivation of the expression for frequency shift may be of value

(see Fig. 191). We consider a radar signal, transmitted at a carrier frequency f_0, impinging on a stationary target. The number of waves passing the target, per second, is f_0. Suppose now that the target moves at velocity v cm per sec toward the radar, that is, it moves a distance v cms in 1 sec. In so doing it passes through several additional waves, each c/f_0 cm long, where c is the velocity of propagation. The target thereby encounters vf_0/c additional waves per sec and the total number of waves impinging on the target is $f_0 + vf_0/c$ per sec. This is, therefore, the frequency of the reradiated wave.

The reradiated wave arises from a moving source, and the effect is as though the target were stationary and the radar

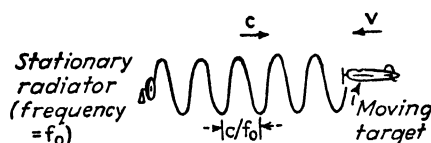



FIG. 191.—Geometry of the Doppler effect.

were in motion. This relative motion gives rise to a second increase in frequency at the radar, of the same magnitude, and by the same mechanism, as the first increase. The frequency of the reflected wave is thus increased in two steps, the first as a result of the motion of the target through additional waves as they impinge on it, and the second as a result of the relative motion of radar receiver toward the target. The frequency of the signal induced in the receiver is, then

$$f_r = f_0\left(1 + \frac{2v}{c}\right) \tag{424}$$

The Doppler shift in frequency Δf is the difference between the transmitter frequency f_0 and the received frequency f_r, that is,

$$\Delta f = f_r - f_0 = \frac{2v}{c} f_0 \tag{425}$$

Similarly, when the target moves away from the radar at a velocity v, the reradiated frequency is lower than the transmitted frequency and the difference between them is given by Eq. (425). The velocity v, it will be noted, is the velocity of the target directly toward or away from the radar. If the motion is

in any other direction, v represents the radial component of the target velocity.

The quantity $2v/c$ in Eq (425) expresses the Doppler shift as a fraction of the carrier frequency. When v is expressed in statute miles per hour, this fraction is $3v \times 10^{-9}$. This is equivalent to 30 cps per 100 mph velocity, per 100 megacycle carrier frequency. Figure 192 gives the magnitude of the Doppler

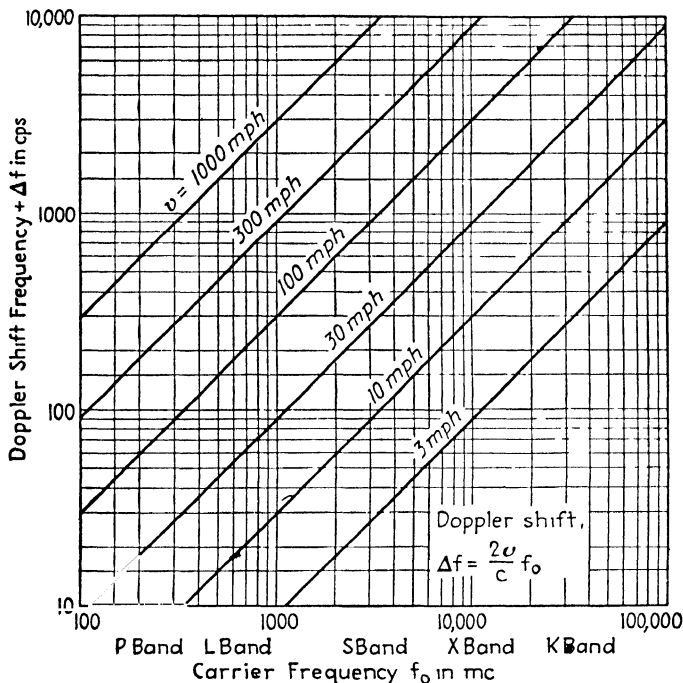



FIG. 192.—Magnitude of frequency shift due to Doppler effect at various frequencies and relative velocities

frequency shift over the range of carrier frequencies from 100 to 100,000 megacycles and radial target velocities from 1 mph to 1,000 mph. The Doppler shift extends from below the a-f range for slow targets observed with l-f carrier signals to well above the audio range for fast targets observed with h-f carriers.

Detection of the Doppler shift may be accomplished by a variety of methods. The transmitted and received frequencies may be mixed in a detector that develops the frequency difference as an a-f beat note. In c-w radars, the beat note appears as an a-f sine wave that may be observed aurally by headphones or

loudspeaker. In pulsed radars, the beat note appears as an envelope that modulates the amplitude of the received pulses (Fig. 193). The envelope may be observed directly on a slow-speed oscilloscope trace, or it may be observed aurally, after passage through an auxiliary detector ("third detector") that recovers the envelope from the pulses.

These methods are limited by the necessity of allowing sufficient time, during the interception of the radar signal by the moving target, for the a-f beat note to build up to appreciable amplitude. Consequently the rate of angular scanning is prohibitively slow for most applications. A more flexible method of detecting the Doppler shift is the so-called "coherent pulse system," in which the radio frequency phase of successively received pulses is compared to reveal any change in phase due to the Doppler shift. Pains are taken to arrange that no such change occurs in echoes arising from stationary objects. By arranging for self-cancellation of signals displaying no phase shift, it is possible to remove from the radar indicator all evidence of stationary reflections, while preserving to a considerable degree the echoes from moving targets.

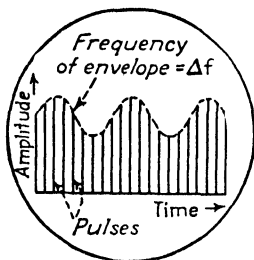



FIG. 193.— Detection of moving target by amplitude modulation of received pulse sequence (slow-sweep type A indicator).

Bibliography

SLATER, J. C.: "Microwave Transmission," McGraw-Hill Book Company, Inc., New York, 1942.

SKILLING, H. H.: "Fundamentals of Electric Waves," John Wiley & Sons, Inc., New York, 1942.

BRAINERD and Others: "Ultra-High-Frequency Techniques," D. Van Nostrand Company, Inc., New York, 1942.

SARBACHER and EDSON: "Hyper and Ultra-High Frequency Engineering," John Wiley & Sons, Inc., New York, 1943.

BARROW and CHU: Theory of the Electromagnetic Horn, *Proc. I.R.E.*, **27,** 51 (January, 1939).

BARROW and GREENE: Rectangular Hollow-Pipe Radiators, *Proc. I.R.E.*, **26,** 1498 (December, 1938).

BARROW and LEWIS: The Sectoral Electromagnetic Horn, *Proc. I.R.E.*, **27,** 41 (January, 1939).

RAMO and WHINNERY: "Fields and Waves in Modern Radio," John Wiley & Sons, Inc., New York, 1944.

PART II

Radar Circuits and Components

CHAPTER V

INTRODUCTION TO RADAR DESIGN

Introduction.—Part II of this book, on which the reader now embarks, is intended to serve as a guide to the design of circuits, structures, and components widely used in radar practice. The design of each element of a radar is governed primarily by the peculiar function it must perform. Since the functions are to a large extent common to all radars, it is convenient to study each element separately, as outlined in the following chapters. As a preliminary, we shall examine briefly the interrelationships between the elements of a radar system and note how the design parameters of one element react on those of other elements.

112. Radar Characteristics and Specifications.—The design of a radar begins with consideration of its intended use, that is, the *function* to be performed by the radar as a whole. The uses generally divide into three categories:

1. Warning and surveillance of activity, including identification
2. Aids to the direction of weapons, that is, gunfire control and searchlight control
3. Observation of terrain echoes or beacons for navigation and control of bombing

While many radars combine several of these functions, they are intended primarily for one use, and they may perform a highly specialized function within any of the general classifications.

After the function of the radar has been stated, it is necessary to consider the *performance requirements.* These specify in general terms the degree of performance required with respect to the intended function of the radar but without reference to the manner of accomplishing the desired end. In military parlance, the performance requirements are known as "military characteristics." The requirements are generally listed in two groups:

1. Physical Requirements. Permissible size, weight, and shape. Type of mounting or vehicle on which used, or in which

transported. If the primary power source is separate, the power requirements are usually stated in this group.

2. Target Data Requirements. Maximum detection range on specified targets at specified aspects. Minimum detection range. Limits of scan in azimuth and elevation. Required accuracy of range, elevation, and azimuth data. Allowable interval between observations, that is, required continuity and allowable error of tracking. Degree of distinction required between multiple targets and between target and background.

The performance requirements in turn lead to *technical specifications*, that is, numerical determinations of the radar system constants. These fall into six categories:

a. Beam Specifications. Angular size, shape, and energy distribution

b. Scanning Specifications. Type of scan, angular rates and limits

c. Radio Frequency. Chosen by reference to the beam requirements in relation to allowable radiator size, target characteristics, and power and sensitivity of available equipment

d. Pulse Specifications. Power, width, and repetition rate

e. Receiver Specifications. Sensitivity, noise figure, bandwidth, and recovery time

f. Indicator Specifications. Types and sizes of indicator tubes and deflection patterns.

The technical specifications may be viewed as quantitative system constants which, if combined in the radar, will permit the performance requirements to be met in a manner suitable to the intended use. These specifications lead to the selection and design of circuits, structures, and components. The final step in the design is the integration of these elements into a prototype equipment from which manufacturing specifications may be drawn. In the following paragraphs, we shall examine in more detail the technical specifications and their interrelations.

113. Technical Specifications. The Radiated Beam.—The aspect of a radar most intimately connected with its performance on a specified target is the radiated beam. Considering the usual symmetrical beam of conical shape, the basic specification is the angular width. The beam width is specified numerically in degrees or mils (1 mil = 0.056 deg) and is measured between the half-power points in the radiation pattern. These are the

limits at which the radiated power falls to one-half the value radiated along the axis of maximum radiation.

The required value of beam width is determined by the desired coverage of the radar in applications involving warning or surveillance or by the required precision of angular data in such applications as gunfire control. Beams as wide as 30 deg. have been used for early warning of distant aircraft, whereas beams as narrow as 0.6 deg are used for precise determinations of the position of small targets. For gunfire control, lobe switching or conical scanning is employed to refine the angular accuracy, resulting in a precision of 0.03 deg in certain gunfire-control radars.

Another property affected by the beam width is the resolution of the system, that is, the ability of the radar to distinguish in angle among several closely spaced targets, as well as to effect maximum contrast between a target and a reflecting background.

Closely associated with the beam width when ground or sea reflections are present is the vertical coverage diagram of the radiated energy. The vertical diagram must be known to establish the lower limit of coverage in elevation, as well as to predict gaps in the coverage. Moreover, the envelope of the lobes in the vertical coverage diagram may be used to give an approximate elevation angle of the target. The vertical coverage diagram generally must be specified by a plot, which may be computed by the methods of Sec. 102.

The beam width specified by reference to the target data requirements must be subjected to scrutiny with respect to several other technical specifications. First, since the beam width enters directly into the radar equation, it affects the maximum detection range. In general, for a given transmitted power and receiver sensitivity, as reference to Eq. (23) shows, the maximum detection range obtainable on a given target is inversely proportional to the beam width. In the second place, the beam width must be examined with respect to the angular scanning rates and pulse repetition rate, by reference to Eq. (1), to ensure that a sufficiently large number of pulses falls on each point of space, during scanning, to reveal a point target. Finally, the required beam width can be obtained, as Eq. (14) shows, only when the area of the radiator covers a specified number of

square wavelengths of the radio frequency used. Hence, the allowable radiator size and choice of radio frequency enter into the determination of beam width. If all these interrelated factors can be satisfied simultaneously, the prescribed value of beam width may be achieved. Otherwise, a compromise is necessary.

In many applications a symmetrical beam is not satisfactory, and a beam of particular shape and energy distribution must be specified. Usually such unsymmetrical beams have two axes of symmetry, each one of which may be considered independently with respect to all the factors enumerated above except wavelength (unless two transmitters, whose frequencies may be chosen independently, are available). The beam width along each of the axes of symmetry is computed by reference to the linear dimension of the radiator relative to the wavelength, by Eq. (14).

114. Scanning Specifications.—The factors governing scanning have already been discussed at some length in Chap. I, (pages 25–31). The first matters to be decided are the type of scan (circular, helical, conical, etc.) and the angular limits. The choice is governed by the application of the radar, that is, by the volume of space to be covered. The rate of scanning, the next question, is decided primarily by the time permissible between successive observations of the target, and by consideration of available scanning mechanisms in relation to the size and weight of the radiator. As stated previously, the scanning rate must be coordinated with the beam width and pulse rate to assure adequate coverage of each point in the space scanned.

Finally, the scanning rate, in conjunction with the scanning pattern, determines the interval between observations, that is, the continuity of tracking. When a wide beam is used, it may be moved slowly to cover a wide area, as in early-warning applications. The required scanning rate assumes intermediate values when the beam width and area to be covered have intermediate values, as they do in radars intended for navigation, bombing, and air-ground or air-sea surveillance. For gunfire control and for tracking small targets, such as shells in flight, the beam must be narrow, and if the space to be covered is substantial the scanning rate must be high. Tracking fast-moving targets, such as rocket aircraft, also requires fast scanning. Equations (1) and (6) in Chap. I govern the selection of scanning rates in various types of scanning.

For many purposes, particularly for gunfire control, automatic tracking of the target is required. This involves special consideration in the choice of scanning methods. Usually, for this purpose, conical scanning (to obtain angular precision and to indicate the sense of the target deviation) is combined with circular or helical scanning (to obtain the required wide coverage of the target area). For other purposes substantially continuous observation of the target area (interval between observations less than 1/10 sec) is required. This is particularly desirable when the target is a very fast-moving object, such as a mortar shell, which may appear anywhere within a large field of view and be visible for but a few seconds. High-speed scanners have been designed for such purposes, scanning a sector of 20 deg in 1/30 sec or less. The design of such equipment is hardly a routine job, but the design problem is nevertheless completely stated in terms of the scanning rates and limits in relation to the beam width and pulse rate.

115. Choice of Radio Frequency.—This matter, like scanning, has received some attention in earlier chapters. A primary determining factor in the choice is the allowable size of the radiator relative to the required angular beam width. Even when these factors allow some freedom of choice, the selection of frequency is not completely a free one. The maximum detection range, with given power and sensitivity on a target of given echo area, varies with the frequency. The range (1) increases with increasing frequency when the antenna size is fixed, (2) decreases with increasing frequency when the beam width is fixed, and (3) remains fixed when the radiator area is chosen so as to keep the beam width proportional to the square root of the wavelength. Moreover, the echo area of the specified target changes with frequency, and there may be a distinct advantage in choosing a particular wavelength, for example, one that is twice the longest dimension of the intended target. This is a practical procedure in the case of certain small targets such as mortar shells and other projectiles. Even when half-wave resonance of this type is not practical, proper choice of frequency may offer a favorable polar diagram in the reradiation pattern of a specific target.

The frequency chosen has an important effect on the vertical coverage diagram. As Eq. (388) shows, the radar height above the reflecting surface, expressed in wavelengths, determines

the separation between the lobes of the vertical diagram. If reflections can be prevented by the use of a narrow beam, the vertical coverage is uniform and it is not affected by frequency. Otherwise, a high frequency is often desirable because the gaps between lobes are then smaller and the coverage is more nearly continuous than if a low frequency is chosen.

The frequency chosen is subject, in the third place, to considerations of available transmitter power and of receiver sensitivity. At present, the power and sensitivity obtainable at all frequencies up to about 3,000 megacycles are about the same, although at any stage of the radar art the lower frequency equipment can be designed for higher average and peak power levels. At frequencies above 3,000 megacycles, however, the available power of existing magnetrons decreases appreciably. The most recent crystal mixers have as low a noise figure at 10,000 megacycles as at 3,000 megacycles, but suffer in this respect when pushed to 30,000 megacycles.

None of these effects is of serious consequence if the target is large enough and the desired maximum range small enough to permit the use of a power level within available limits. But in most radar design, the specified target grows ever smaller and the required maximum range ever larger as the development proceeds. Thus it is seldom safe to assume that plenty of power is available and to choose a high frequency simply on that account.

The frequency chosen must also be examined with respect to anomalous propagation effects, attenuation and absorption due to rainfall and cloud formations, and absorption in the atmosphere (important only at very short wavelengths). Finally, during radar warfare, the relationship of the frequency with the enemy countermeasure activity must be considered. High frequencies, being less fully developed than the low, are generally advantageous in this respect. Table IV (page 135) gives the frequency bands in common use.

116. Pulse Specifications.—The effect of the pulse specifications on the design of a radar is comprehensive. As indicated in Chap. I (page 57), the basic specification of a pulse respecting the maximum range on a given target is its *energy*. The energy is the product of the pulse width by its maximum (peak) power, or the quotient of the average power by the pulse rate. As the energy of the pulse is increased, the maximum detection

range on a given target increases, provided, of course, that the system is designed for optimum signal-to-noise ratio. Thus, no increase in range is achieved by increasing the pulse width unless the bandwidth of the receiving system is proportionately narrowed. Nor is any effect produced if the peak power is increased at the expense of narrowing the pulse proportionately since the pulse energy then remains constant.

These relationships indicate clearly that the specification of the pulse width, peak power, average power, and pulse rate must be coordinated with other factors in the design, particularly with the receiver bandpass characteristic. Each item must also be chosen with respect to other performance requirements and technical specifications, outlined immediately below.

The *pulse width*, as we have seen in Chap. I, determines the minimum detection range. More accurately, the minimum detection range corresponds to the time at which the receiver recovers from the effects of the transmitted pulse. The recovery time of the t-r (duplex) system and of the succeeding i-f (intermediate-frequency) and video amplifiers in the receiver may limit the minimum range to a value well beyond that set by the pulse width. The pulse width sets the lower limit, and this limit may be approached within a few microseconds in properly designed receiver systems.

The pulse width also determines the resolution of the radar in the range coordinate, and this is often an important matter. Thus, if closely spaced targets are to be distinguished from one another, or if a target is to be resolved against a background, a short pulse is advantageous.

The pulse width must also be considered in relation to the pulse rate to assure that the average power required is within the operation limits of available r-f (radio-frequency) generators, and this is particularly true when high power is required from h-f (high-frequency) cavity-magnetron generators.

These considerations have led to a well defined range of values, from about 0.25 μsec (used in h-f radars of high resolving power) to about 20 μsec [used in l-f (low-frequency) low-pulse-rate, long-range-warning radars.] A general purpose figure is 1 μsec, the width adopted in the majority of 10-cm and 3-cm microwave equipment.

The *pulse rate*, in addition to effects already noted in connection with average power and pulse energy, determines the maximum

detection range. This is not to say that targets are invisible beyond a certain range if the pulse rate is too high. It means merely that an ambiguity concerning the range of the target may exist unless the pulse interval is long enough to allow the most distant echo to arrive before the next pulse is sent out. Since such ambiguity is nearly always fatal to the tactical use of the radar, the maximum value of the pulse rate is in effect predetermined by the maximum detection range to be expected or required under normal propagation conditions.

The pulse rate should not, on the other hand, be lower than that demanded by the prescribed maximum range, since a low pulse rate limits the angular scanning rates to correspondingly low values as we have seen in Sec. 114. But if low angular rates are permissible (when a wide beam is used to cover a narrow sector), a low pulse rate suffices and has the advantage of permitting the production of a high-energy pulse from a low-average-power source. Pulse rates vary from 100 per second to 10,000 per second depending on the application.

The *pulse shape* has not received detailed attention previously except in respect to width and peak amplitude. Generally, the pulse shape, as indicated by the envelope of the transmitted wave, is nearly rectangular. In microwave equipment, the need for a rectangular pulse arises from the peculiar characteristics of the cavity magnetron, which tends to change its frequency of operation as the modulating current fed to it changes. A rectangular wave is thus desirable in the interest of preserving constant frequency during the pulse. Likewise, in lower frequency radar equipment, the rectangular waveform is used because it stabilizes transmitter operation.

The rectangular wave is undesirable in that it does not conserve ether space. The receiver bandwidth is generally taken as that which will produce a maximum amplitude of response in the output, relative to the noise, and this bandwidth takes in only the first zero of the rectangular pulse spectrum. The energy transmitted outside this region is wasted and may cause interference with other equipment. But up to the present it has not proved feasible to restrict the transmitted sidebands in the interest of economy or mitigation of interference.

The peak power, the final specification of the pulse, has already been mentioned in connection with energy and average power

considerations. These, rather than the peak power itself, determine the maximum range. Another aspect of the peak power is the ability of the transmission system (waveguides or coaxial lines) to withstand voltage stresses. It is economical to use a transmission system whose dimensions do not greatly exceed the minimum values set by sparkover when the line is properly terminated and to depend upon proper adjustment of the system to avoid breakdown due to standing waves. Finally, the peak power must be chosen with reference to the ability of the t-r system to reduce the power transmitted to the receiver to a safe value. If high power is necessary, special precautions must be taken in the t-r system (see Chap. VII). Peak power values from 1 kw to 2,000 kw are employed in practice.

117. Receiver Specifications.—The basic receiver specifications, that is, those related to the other parts of the system, are the sensitivity, noise figure, bandwidth, and recovery time. Every radar receiver must be designed with sufficient sensitivity to saturate the output with the noise existing across the input at full gain. When this condition is satisfied, the maximum detection range of the system is set by the noise introduced by the receiver relative to $kT\Delta f$ (noise figure, Chap. II, page 133), and this in turn involves the bandwidth Δf. The bandwidth is chosen, for most purposes, at $1/d$ to $2/d$ megacycles where d is the pulse width in microseconds. For special applications, the bandwidth may be substantially wider than this, particularly when high resolution is required in the range coordinate, and the pulse width delivered to the indicator must not be substantially wider than that received from the target. Bandwidths of from 0.5 megacycles to 20 megacycles have been employed in radar receivers. The general-purpose figures vary from 1 to 2 megacycles. Aside from the provision of bandwidth to produce a pulse of maximum amplitude, at specified width, bandwidth must also be provided to accommodate incidental variations in the transmitter and local oscillator frequencies.

118. Indicator Specifications.—The indicator specifications involve the choice of type of indication and indicator tubes (size and type of phosphor). The type of indication adopted is closely related to the type of scanning used, on the one hand, and to the end-use of the target information, on the other.

Type A indication is embodied in nearly every radar for purposes

of tuning and otherwise adjusting the radar to highest efficiency. Except in the early radars, type A is seldom used to provide target information as such because it does not reveal angular coordinates. The most employed indication types for immediate observation of the target are type B (or its variant, type E) and the plan-position indicator. The choice between them rests on the intended use. The plan-position indicator is particularly useful for airborne navigation and bombing because the terrain and target echoes retain their shapes as they are passed by. The type B is used where greater resolution of angular information is required at close range, or where the electromechanical complexities of a plan-position indicator deflecting system are not deemed worth while.

When the type of indication is decided (and generally more than one type is needed), certain numerical aspects must be specified. These involve the range scales (generally two or more maximum range limits are desired, with or without accompanying change in pulse rate), and the angular limits, which usually coincide with the angular limits of the scanning. Such matters as the size of the indicator tube are usually determined by such physical requirements as the allowable size and weight. The brightness and contrast of the phosphor may be important in applications where the indicator screen cannot be shielded from light. The persistence of the phosphor must also be considered in relation to the pulse rate and the number of pulses transmitted while the beam passes over a point target. If the latter number is small, lack of sensitivity may result from incomplete integration of the signal on the phosphor surface. In special applications, of course, a long-persistence phosphor may not be desired, especially when the entire scanning sequence repeats itself several times a second.

119. Technical Specifications of Representative Military Radars.—It is evident that a good deal of judgment and experience must go into the selection of the specifications just described. As an indication of present design trends, and as an aid to the reader in orienting himself, Table IX lists the technical specifications of a number of representative radars.

120. Selection of Circuits and Components.—The circuits and components required to satisfy the technical specifications of a radar may be studied in several distinct groups. The most

convenient breakdown is in accordance with the functional elements shown in Fig. 194. The operation of the radar begins with the *timer*, which establishes the pulse-repetition rate and coordinates the timing of the other elements of the radar. The timer controls the *modulator* that provides a modulating pulse of the requisite power and width, and of proper shape, to actuate the *r-f oscillator* or *transmitter*. The transmitter generates r-f pulses, within the envelope established by the modulator, and

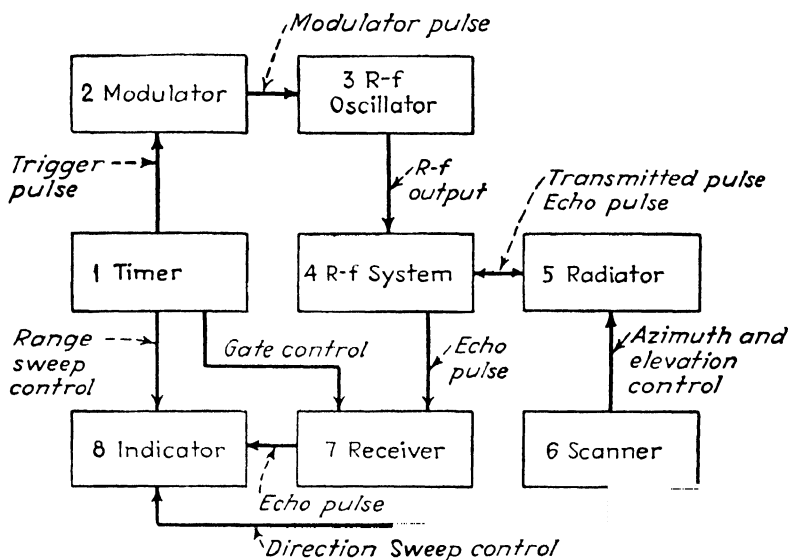



FIG. 194.—Essential elements and interconnections of a radar system.

applies the pulses to the *r-f system*, that is the transmission and duplex switching system intermediate to the transmitter, radiator, and receiver. The *radiator* forms the radiated beam and presents the absorption area for reception of the echoes. The *scanner* is an electromechanical device that generates the angular motion of the beam through space, in accordance with the established scanning sequence, and generates electrical signals to represent the angular coordinates of the beam. The *receiver* accepts the echo pulses from the r-f system, converts to and amplifies at intermediate frequency, and detects and further amplifies at video frequencies. In many forms of radar the timer actuates the sensitivity control of the receiver so that it operates only during the period in which echoes are to

TABLE IX.—TECHNICAL SPECIFICATIONS OF REPRESENTATIVE UNITED STATES RADARS

Type no.	Function	Size and weight	Wavelength frequency	Radiator size beam width	Scanning specifications
			Ground-based Equipment		
SCR-270-B	Ground-based, long-range, early warning against aircraft	Three large trucks and a trailer, 81,400 lb	270–290 cm 104–112 mc	4 dipoles wide, 9 dipoles high, 28° in azimuth, 10° in elevation	Circular through 360° in azimuth only at 1 rpm
SCR-602-A	Transportable, light weight, early warning against aircraft	300 cu ft, 4,709 lb	140 cm 212 mc	Four Yagis, on 6 ft centers, 24° in azimuth, 30° in elevation	Circular through 360° at 6 rpm in azimuth only
SCR-584	Mobile search and track for antiaircraft gunfire control	8 by 10 by 20 ft, trailer, 20,000 lb	10–11 cm 2,700–2,900 mc	6-ft paraboloid, 4° in azimuth and elevation, 7° with conical scanning	Helical through 360° at 6 rpm, elevation from −5° to +89°. Conical scan for automatic tracking
AN/CPS-1	Fixed microwave long-range early warning and fighter direction	25 ft radiator on 40 ft platform. About 100,000 lb	10–11 cm 2,700–2,900 mc	Two 25 ft parabolic cylinders back to back, 0.8° in azimuth; 23° in elevation	Circular at 4 rpm maximum. Elevation variable from 0° to 4°
AN/MPG-1	Mobile search and track for coastal gunfire control	8 by 10 by 20 ft trailer, 28,000 lb	3 cm 10,000 mc	Rapid-scan folded horn and parabolic cylinder. Azimuth 0.6°; elevation 3°	Rapid scan over 10° sector at 16 scans per sec. Circular scan through 360° for search

TABLE IX.—TECHNICAL SPECIFICATIONS OF REPRESENTATIVE UNITED STATES RADARS.—(*Continued*)

Type no.	Pulse specifications	Receiver specifications	Indicator specifications	Maximum range
SCR-270-B	Width 10–30 μsec Rate 621 pps Peak power 100 kw Energy 1–3 joules	Noise figure 12 db Bandwidth 1.5 mc	Type A, 5 in. Type ppi, 12 in. (Type 270-BB only)	80–120 miles on bombers 50–75 miles on fighters
SCR-602-A	Width 2 μsec Rate 400 pps Peak power 100 kw Energy 0.2 joules	Noise figure 15 db Bandwidth 1.75 mc	Type A, 5 in. Type ppi, 9 in.	45 miles on bombers 40 miles on fighters
SCR-584	Width 0.8 μsec Rate 1,707 pps Peak power 300 kw Energy 0.24 joules	Noise figure 15 db Bandwidth 1.7 mc	Two type J, 3 in. Type ppi, 7 in.	34 miles on bombers Range accuracy ±15 yd Angular accuracy ±0.06°.
AN/CPS-1	Width 1 μsec Rate 350 pps Peak power 700 kw Energy 0.7 joules	Noise figure 15 db Bandwidth 1.7–2.0 mc	Five type B 12 in. ppi Five type B 7 in. One type A 5 in.	175 miles on single bomber 100 miles on single fighter
AN/MPG-1	Width 1 and 0.25 μsec Rate 1,024 and 4,097 pps Peak power 60 kw Energy 0.06 and 0.015 joules	Noise figure 17 db Bandwidth 10 mc	Two type B, 7 in. One ppi, 7 in.	28 miles on battleships 9 miles on surfaced submarine

TABLE IX.—TECHNICAL SPECIFICATIONS OF REPRESENTATIVE UNITED STATES RADARS.—(*Continued*)

Type no.	Function	Size and weight	Wavelength frequency	Radiator size beam width	Scanning specifications
Airborne Equipment					
ASB	Airborne search for marine surface targets	120 lb	58 cm 515 megacycles	Two Yagi arrays, lobe switched. 60° in azimuth.	Manual adjustment through 90°
SCR-720	Airborne search and interception	415 lb	10 cm 3,000 megacycles	29-in. paraboloid. 10° in azimuth and elevation	Helical at 100 or 360 rpm. 10°, 15°, or 30° in elevation
AN/APQ-13	Airborne radar-controlled bombing	619 lb	3 cm 10,000 megacycles	30-in. composite reflector, spheroid plus paraboloid. 3° in azimuth. Csc[2] coverage in elevation	Circular at 20 rpm. Elevation tilt +10° to −45°
AN/APS-3	Airborne search and interception	235 lb	3 cm 10,000 megacycles	17-in. paraboloid. 5° in azimuth and elevation	Sector scan over 160° at 35 scans per min. Elevation scan over 2°. Tilt over 24°.
AN/APQ-7	Airborne radar-controlled bombing	775 lb	3 cm 10,000 megacycles	16-ft dipole array in auxiliary wing 0.4° in azimuth; 30° in elevation	Sector scan over 60° in azimuth by phase shift at 1.5 scans per sec

TABLE IX.—TECHNICAL SPECIFICATIONS OF REPRESENTATIVE UNITED STATES RADARS.—(*Continued*)

Type no.	Pulse specifications	Receiver specifications	Indicator specifications	Maximum range
ASB	Width 2 μsec Rate 400 pps Peak power 5–10 kw Energy 0.01 joules	Noise figure 12–20 db Bandwidth 1.4–3 mc	Type L, 5 in.	40 miles on battleships 6 miles on bombers
SCR-720	Width 0.75 μsec Rate 1,500 pps Peak power 100–150 kw Energy 0.08–0.11 joules	Noise figure 11–15 db Bandwidth 3 mc	Two type B, 3 in. and 5 in. One type C, 5 in.	10 miles on bombers 5 miles on fighters
AN/APQ-13	Width 0.5–1.1 μsec Rate 624–1,348 pps Peak power 35 kw Energy 0.018 joules	Noise figure 17 db Bandwidth 5 mc	Two type ppi, 5 in.	100 miles on ground targets 50 miles on battleships 5 miles on bombers
AN/APS-3	Width 1.0 μsec Rate 700 pps Peak power 35 kw Energy 0.035 joules	Noise figure 18 db Bandwidth 2.5–3 mc	Two type B, 5 in.	50 miles on battleships 8 miles on bombers
AN/APQ-7	Width 0.4–1.0 μsec Rate 400, 800, 1,600 pps Peak power 50 kw Energy 0.05 joules	Noise figure 15 db Bandwidth 4 mc	Two sector type ppi showing true ground range, 5 in Type A 3 in.	75–100 miles on cities

TABLE IX.—TECHNICAL SPECIFICATIONS OF REPRESENTATIVE UNITED STATES RADARS.—(*Continued*)

Type no.	Function	Size and weight	Wavelength frequency	Radiator size beam width	Scanning specifications
			Shipborne Equipment		
SK	Shipborne search for aircraft targets	4,600 lb	150–155 cm 193–198 mc	Mattress of 36 horizontal dipoles 20° in azimuth and elevation.	Circular through 360° at 5 rpm. Elevation fixed
SL	Shipborne search for surface targets	1,500 lb	10 1–10 4 cm 2,915–2,967 mc	Truncated paraboloid 20 by 42 in. 6° in azimuth; 12° in elevation	Circular through 360° at 18 rpm
SO-3	Shipborne search for surface targets	750 lb	3 cm 10,000 mc	Truncated paraboloid 8 by 24 in. 3° in azimuth; 9° in elevation	Circular through 360° at 5 rpm
Mark 4	Shipborne control of gunfire	3,900 lb	43 cm 700 mc	6 by 7 ft. Two parabolic cylinders, fed by 8 dipoles each. 12° in azimuth and elevation	Four-way lobe switching at 30 per sec
Mark 13 Mod 0	Shipborne control of gunfire	6,000 lb	3 3 cm 9,000 mc	Truncated paraboloid 8 by 2 ft. 0.9° in azimuth; 3.5° in elevation	Sinusoidal high-speed scan over 10° sector at 5 scans per sec

TABLE IX.—TECHNICAL SPECIFICATIONS OF REPRESENTATIVE UNITED STATES RADARS.—(*Continued*)

Type no.	Pulse specifications	Receiver specifications	Incicator specifications	Maximum range
SK	Width 5 μsecs Rate 60 pps Peak power 200 kw Energy 1.0 joules	Noise figure 6 db Bandwidth 1.5 mc	Type A, 5 in. Type ppi, 12 in., plus ppi repeaters	110 miles on bombers 20 miles on battleships
SL	Width 1.5 μsec Rate 800 pps Peak power 120 kw Energy 0.18 joules	Noise figure 13 db Bandwidth 2 mc	Type ppi, 7 in.	15 miles on bombers 20 miles on battleships
SO-3	Width 1.0 μsec Rate 400 pps Peak power 20 kw Energy 0.02 joules	Noise figure 15 db Bandwidth 2 mc	Type ppi, 5 in.	10 miles on bombers 20 miles on battleships
Mark 4	Width 1.5 μsec Rate 1639.3 pps Peak power 20 kw Energy 0.03 joules	Noise figure 10 db Bandwidth 1 mc	Two type A, 3 in. One type A, 5 in.	23 miles on bombers 17 miles on battleships
Mark 13 Mod 0	Width 0.3 μsec Rate 1,800 pps Peak power 34–45 kw Energy 0.01 joules	Noise figure 15 db Bandwidth 8 mc	Six type B, 3 in.	17 miles on bombers 23 miles on battleships

be received. The output of the receiver is applied to the *indicator*, which consists of several (two or more) cathode-ray tubes and auxiliary-deflection circuits operating under the control of the timer and the scanner.

The design of a radar involves the selection of suitable circuits and components to fulfill the eight basic functions listed above and to meet at the same time the physical requirements and technical specifications previously discussed. If the design is routine, the selection may be made by reference to common usage and previous designs. Preferably, the choice is made on a more fundamental basis, that is, by study of the comparative merits of alternative circuits or structures. The latter approach is clearly preferable if the design is to show substantial improvement over previous work.

In the two following chapters, separate consideration is given to basic aspects common to all radar practice: pulse circuit design and r-f structure design. From among the many alternatives presented, the designer can select the circuits and structures best fitted to particular needs.

Bibliography

(Descriptions of Radar Equipment)

COLTON, R. B.: Radar in the U. S. Army (SCR-268 and SCR-270), *Proc. I.R.E.*, **33,** 740 (November, 1945).

BERKNER, L. V.: Naval airborne radar, *Proc. I.R.E.*, **34,** 671 (September, 1946).

BYRNES, I. F.: Merchant marine radar, *RCA Rev.*, **7,** 54 (March, 1946).

WATSON, C. W.: Ground-controlled approach for aircraft (type AN/MPN-1), *Electronics*, **18** (11) 112 (November, 1945).

STRAUS, H. A., and others: Fire-control radar AN/MPG-1, *Electronics*, **18** (12) 92 (December, 1945); **19** (1) 110 (January, 1946); **19** (3) 140 (March, 1946).

FINK, D. G.: SCR-268 radar, *Electronics*, **18** (9) 100 (September, 1945).

———: Radar specifications, *Electronics*, **18** (11) 116 (November, 1945).

———: SCR-584 Radar, *Electronics*, **18** (11) 104 (November, 1945); **18** (12) 104 (December, 1945); **19** (2) 110 (February, 1946).

-: F-m radar altimeter, *Electronics*, **19** (4) 130 (April, 1946).

ZAHL AND MARCHETTI: Radar on 50 centimeters (AN/TPS-3), *Electronics* **19** (1) 98 (January, 1946); **19** (2) 98 (February, 1946).

HOLDAM, MCGRATH, AND COLE: Radar for blind bombing (AN/APQ-13 and AN/APS-15) *Electronics*, **19** (5) 136 (May, 1946); **19** (6) 142 (June, 1946).

BARNES, C. B.: Radar for carrier-based planes (AN/APS-4), *Electronics*, **19** (10) 100 (October, 1946).

CHAPTER VI

BASIC PULSE CIRCUITS

In this chapter we examine the operation of pulse circuits used in radar systems. There are two classes of such circuits: (1) those concerned with the generation and amplification of the transmitted and received pulses and (2) those used to coordinate the operation of the receiver and indicator. Both types of circuit employ similar principles. For example, a multivibrator may be used to generate the basic pulse sequence. The same circuit, with different design constants, may produce broad rectangular waveforms, synchronized with the basic pulse sequence, to control the gain of the receiver or the brightness of the cathode-ray beam of the indicator. Since a few basic circuits may thus be used to perform various functions in a radar system, it is convenient to study the operation of these circuits first. Thereafter it is simple to identify them in particular applications and to understand their operation without further inquiry.

121. Function of Pulse Circuits.—The basic functions of pulse circuits are as follows:

1. Establishment of the basic pulse rate
2. Direct (unsynchronized) generation of pulses
3. Synchronized production of pulses
4. Synchronized production of extended waveforms
5. Modification of pulse shape
6. Modification of the repetition rate (frequency multiplication or division)
7. Amplification of current, voltage, power, or impedance level
8. Introduction of time delay between two or more sequences of pulses or waveforms

The above functions are performed by circuits discussed in the following sections. Many of the circuits are capable of performing more than one function simultaneously, and there

are, of course, many possible variations in each type. In what follows, space is available to describe only the basic circuit and one or two of the more important variations.

122. Basic Timing Sources.—The basic timing sources used in radar are oscillators and mechanically driven interrupters. Two types of oscillator are used: the sinusoidal type and the relaxation type.

The sinusoidal oscillators used are generally familiar to technicians in the radio field. When high precision of the timing is required, piezoelectric (quartz-crystal) oscillators are used in conjunction with frequency-dividing circuits. The resistance-capacitance sine-wave oscillator (Wien bridge) also serves as a stable source of timing, and it has the advantage of a wide frequency range. The third type of sine-wave oscillator employs the conventional LC circuit in any of the conventional forms (Hartley, Colpitts, electron coupled, etc.). The operation of these oscillators is well known. The choice of the sine-wave oscillator is dictated by the need for precise timing, as mentioned above, as well as by the simplicity and precision with which the timing reference may be shifted by conventional sinusoidal phase-shifting circuits (see Sec. 141).

The two forms of relaxation oscillator most widely used are the multivibrator and the blocking oscillator. These circuits are also well known in the radio art. They do not possess the high frequency stability of the sine-wave oscillators, but they can generate pulses or (in the case of the multivibrator) extended waveforms as they oscillate. The output of a sine-wave oscillator must, on the other hand, be modified to produce pulses or extended waves, and hence sine-wave systems generally require more tubes than relaxation systems. The choice is essentially one of precision vs. simplicity of circuit design.

The third type of basic timing source is mechanical in nature. The most widely used mechanical timing device is the motor-driven rotary spark gap. The basic timing is established by the rotation rate of the motor and the number of spark electrodes. The frequency stability of mechanical timers is generally lower than that of electronic oscillators. But the spark gap can generate high-power pulses directly, without the need of successive shaping and amplifying circuits.

123. Direct (Unsynchronized) Generation of Pulses and Extended Waveforms. a. The Multivibrator.—As has just been noted, the multivibrator generates pulses or extended waveforms as it oscillates. The circuit is capable, also, of being synchronized by an external source. When so synchronized the multivibrator can multiply or divide the synchronization frequency and, in conjunction with other circuits, introduce time delay. For these reasons, the multivibrator is one of the most widely applied pulse circuits and its action is worthy of detailed examination.

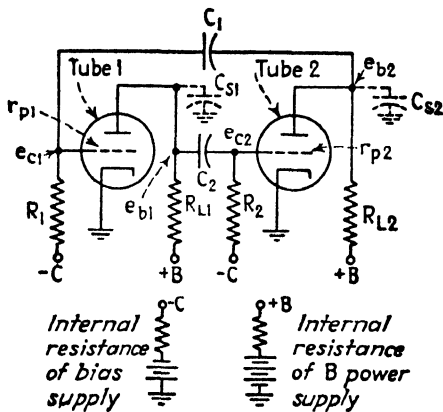



FIG. 195.—The multivibrator circuit.

The multivibrator is essentially two stages of resistance-capacitance-coupled amplification, with the output fed back to the input, as shown in Fig. 195. When the plate potential is first applied, both tubes draw approximately equal plate current. The circuit is then in a condition of unstable equilibrium since any change in the voltage on either grid is amplified and fed back in the same polarity to the same grid.

Assume initially that a casual disturbance causes the grid voltage of tube 1 (Fig. 195) to decrease. Tube 1 amplifies this change and reverses its polarity, applying an increasing signal to the grid of tube 2. This tube amplifies it further, reverses the polarity once more, and applies a large decrease in voltage to the grid of tube 1. The initial decrease of voltage on the grid of tube 1 is thus enhanced cumulatively at a rapid rate, and the grid of tube 2 is driven at once to a voltage value well below cutoff. Tube 1 is then nonconducting and tube 2 conducting.

After the sudden lowering of potential of grid 1, the coupling capacitor C_1 remains charged and this charge holds grid 1 below cutoff. So long as C_1 retains its charge, the circuit remains quiescent. But C_1 discharges through the associated series resistance, composed of R_1, R_{L2}, and r_{p2} in shunt, and the internal impedances of the plate and bias power supplies. Since R_1 is generally the largest resistance, the other elements may be neglected, and C_1 may be considered to discharge in accordance with the time constant R_1C_1.

As the potential of grid 1 rises above the cutoff level, tube 1 begins to conduct, the potential of plate 1 begins to fall, and the potential of grid 2 begins to fall. This is the condition assumed initially, except that the tubes have now exchanged roles. The decrease in voltage of grid 2 is fed back cumulatively in the same polarity. Tube 2 is thus driven to cut off almost instantaneously, and tube 1 assumes full conduction.

The action of the circuit thus consists of rapid shifts of plate and grid potential, as one tube ceases to conduct and the other assumes full conduction. The potential shifts are separated by quiescent intervals during which the grid of the nonconducting tube recovers from the cutoff condition. The total period of oscillation is determined by the sum of the two time constants R_1C_1 and R_2C_2.

The waveforms of interest are those appearing at the plates and grids of the tubes. Let us assume that the circuit is symmetrical, that is, it employs identical tubes and circuit elements in each half of the circuit. Then the waveforms appearing across each tube are as shown in Fig. 196*a*. The waveforms of tube 1 are formed as follows: at time t_0 the grid potential rapidly falls below the cutoff level. Thereafter it recovers along the R_1C_1 exponential curve. Meanwhile e_{b1}, the plate potential of the same (nonconducting) tube, assumes the full voltage of the plate supply and remains at this level until time t_1, at which time e_{c1} has risen to the cutoff level. At this instant e_{c1} is suddenly driven positive and rises above zero potential momentarily as indicated by the small "spike" above the zero axis. While in the positive region grid 1 draws current that charges the associated coupling capacitor. As C_1 is charged, e_{c1} decreases to zero and there remains until time t_2. Between times t_1 and t_2, e_{b1} remains at a low value, corresponding to the plate voltage

at zero grid potential. The spike at time t_1 is reproduced in the plate voltage as shown. At time t_2 the circuit resumes its initial condition (as at time t_0) and the cycle repeats. The waveforms of the opposite tube, e_{c2} and e_{b2} against time, are identical but displaced in time by one-half cycle. The plate waveform is very nearly rectangular in shape, except for the small spike due to grid circuit, whereas the grid waveform comprises an exponential portion and a rectangular portion.

The waveforms in Fig. 196*a* indicate a substantially instantaneous change of voltage at the leading edge of the wave.

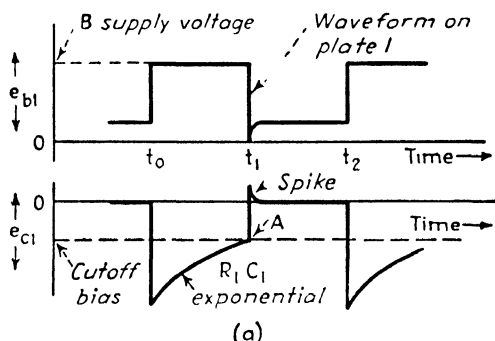



FIG. 196*a*.—Plate and grid waveforms of the symmetrical multivibrator.

Actually the leading edge follows an exponential curve of very short time constant. The grid potential of tube 1 and the plate potential of tube 2 can decrease only as the shunt capacitance C_{s2} is discharged by the plate current of tube 2, through its internal plate resistance r_{p2}. The time constant governing the trailing edge is thus $r_{p2}C_{s2}$. In a typical multivibrator, designed to produce sharp leading edges, C_{s2} may be 50 $\mu\mu$f and r_{p2} 5,000 ohms. Then $r_{p2}C_{s2}$ is 0.25 μsec. The leading edge of the wave, formed as the tube becomes nonconducting, is determined by the charge of C_{s2} through the plate load resistance R_{L2}. The time constant is $R_{L2}C_{s2}$, which may be a fraction of a microsecond.

Fig. 196*b* shows in expanded form the exponential curves and associated time constants of each portion of the wave. It is evident that by proper choice of R_1, C_1, C_{s1}, R_{L1}, r_{p1}, and the corresponding values in the opposite tube, a wide variety of waveforms may be produced.

The detailed computation of multivibrator performance by reference to tube characteristics and circuit elements is difficult,

particularly since it involves the grid-current characteristics of the tubes. A simplified method of computation has been devised by Kiebert and Inglis.[1]

The successive halves of the waveform may be made of unequal width by choosing one coupling time constant shorter

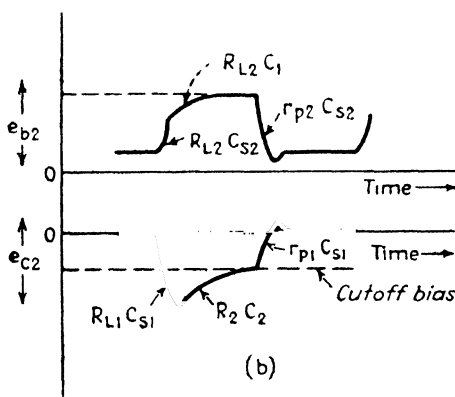



FIG. 196*b*.—Time constants associated with multivibrator waveform.

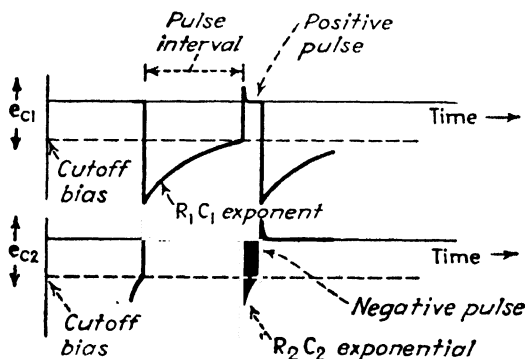



FIG. 197.—Pulses produced by an unsymmetrical multivibrator.

than the other. Suppose that R_2 and C_2 are small compared with R_1 and C_1. Then, as shown in Fig. 197, the conducting period of tube 1 is correspondingly shorter than that of tube 2. In effect, the shorter waves are pulses and the longer ones are pulse intervals. The leading and trailing edges of the pulse are governed by the shunt time constants previously mentioned ($r_{p2}C_{s2}$ and $R_{L2}C_{s2}$) and the width of the pulse by the smaller

[1] KIEBERT and INGLIS, Multivibrator Circuits, *Proc. I.R.E.*, **33**, 534 (August, 1945).

of the coupling time constants R_2C_2. The total period of the wave is governed, to a first approximation, by the longer coupling time constant R_1C_1. Substantially independent control is afforded over the pulse width and the pulse rate by adjusting of R_1 and R_2, respectively.

The multivibrator circuit, flexible as it is, has limitations. The frequency stability suffers from a number of causes. Changes in the circuit elements, induced by temperature and humidity effects, as well as voltage variations and aging of tubes, react directly on the frequency of the output. One of the most serious causes of instability is the variation in the time of recovery of the grid of the nonconducting tube. This tube recovers when the exponential grid potential curve intersects the cutoff bias level (at point A in Fig. 196*a*). If the exponential discharge has by that time nearly completed itself, the intersection of the recovery curve with the cutoff level is not sharp, and the precise point of intersection is readily shifted by changes in contact potential and other casual variations. To improve the frequency stability, it is customary to return the grids of both tubes to a positive bias potential, which aids in discharging the coupling capacitor and hence causes the grid recovery curve to intersect the cutoff level more sharply. The frequency of operation is also increased. Adjustment of the positive bias serves as a convenient control of the operating frequency.

It should be noted that we have treated thus far only the unsynchronized form of multivibrator. As we shall see later, the action of the circuit may be inhibited by negative grid bias so that oscillations occur only when an external synchronizing voltage is applied.

124. b. The Blocking Oscillator.—The blocking oscillator is a device suitable for the direct generation of pulses without external synchronization. The basic circuit is shown in Fig. 198. Essentially, the circuit is a transformer coupled oscillator with the grid isolated from ground by the blocking capacitor C_g. The transformer is so connected that a fall in potential of the plate terminal induces a rise in potential of the grid terminal. Hence, any change in potential at the grid is amplified and fed back to the grid in the same polarity, as in the multivibrator.

To follow the action of the circuit, assume initially that the plate current is cut off by negative grid voltage arising from

the charge on the capacitor C_g. As this capacitor discharges through the associated series resistor R_g to ground, the grid potential rises above the cutoff level, and the tube begins to draw plate current. The plate current, acting on the transformer, induces a further positive change on the grid, and thus the grid is driven rapidly positive, reaching the grid-current region as fast as the transformer will react.

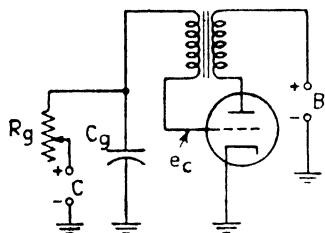



FIG. 198.—The blocking oscillator circuit.

When the grid begins to draw current, the capacitor C_g is charged and the grid voltage reaches equilibrium at zero potential. This leveling off of grid voltage arrests the increase in plate current and the transformer reverses polarity, as if forming the first cycle of an oscillatory wave. The reversal of polarity drives the grid negative, while the charge remains on C_g. Hence, the grid voltage of the tube is reduced below cutoff and is maintained at this level by the voltage across C_g. No further plate current flows; therefore the incipient oscillations in the transformer die out before the first cycle is completed, and the circuit remains at rest in the condition initially assumed above. Thereafter, as C_g discharges, the grid once more rises above the cutoff level and the cycle of conduction is repeated.

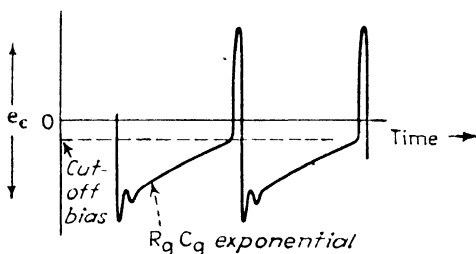



FIG. 199.—Grid waveform of the blocking oscillator.

The waveform of the circuit, shown in Fig. 199, consists of sharp pulses, formed during the conduction period, with longer intervals between them. Negative pulses appear at the plate, positive at the grid. The length of the intervals is determined by the time constant R_gC_g. Since the pulse is, in effect, the first half-cycle of the oscillation of the circuit, its width is determined

by the period at which the circuit would resonate if the blocking capacitor were not present. This self-oscillation frequency is determined by the inductance and distributed capacitance of the transformer windings, both of which should be of low value to produce h-f oscillations, and consequently, narrow pulses.

In addition to this resonance requirement, the transformer must display high losses and close coupling. The close coupling permits the circuit to attain a high amplitude of oscillation on the first cycle, whereas the high losses permit the circuit to cease oscillation abruptly when the plate current is cut off.

As in the case of the multivibrator, the frequency stability depends on the intersection of the grid-voltage recovery curve with the cutoff bias level. The intersection may be sharpened by returning the discharge resistor R_g to the positive terminal of the *B* supply and thus hastening the discharge. The pulse rate may be adjusted by varying either the values of R_g or the grid-bias voltage. The blocking oscillator may be synchronized by applying a positive control voltage to the grid, so that it raises the grid to the conduction point. The synchronized blocking oscillator is useful as a frequency divider, as shown in Sec. 133.

125. Synchronized Production of Pulses from Extended Waveforms. a. Deformation of Sine Waves by Limiters.—An important function performed by pulse circuits is the production of a pulse sequence in synchronism with the leading or trailing edges of more extended waveforms. This is clearly an essential process when pulses are to be formed from a sinusoidal timing source. The process is used, also, to form pulses from the trailing edges of rectangular waves for the purpose of introducing time delay.

To generate a series of precisely timed pulses from extended waves it is essential that the leading or trailing edge (whichever is used as the timing reference) of the extended wave be extremely sharp. The sinusoidal wave is evidently at a disadvantage in this respect because the time of rise of the wave is inherently one-half the full period of the wave. Consequently, if the sine wave is used to control another circuit at a particular level of voltage, the intersection of the sine wave with the control level is not precise.

To overcome this limitation, when sine-wave timing is used, it is customary to deform the sine wave into a rectangular or trapezoidal shape by passing it through limiter amplifier stages

(Fig. 200). As we have seen in Chap. II (Secs. 40 and 41), limiting is accomplished by applying a signal whose amplitude exceeds the grid-voltage cutoff limit in the negative direction or the grid-current limit in the positive direction.

The essential requirement of sine-wave deformation is sufficient amplitude of the applied sine wave to exceed the limits just

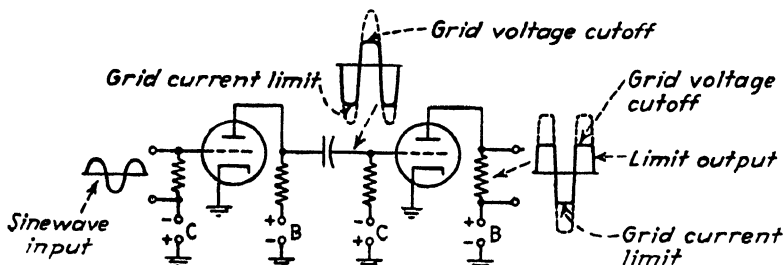



FIG. 200.—Limiting amplifier.

mentioned. If the sine-wave source does not provide sufficient output for this purpose, an auxiliary amplifier must precede the limiter circuit. The gain required in this amplifier may be calculated in terms of the required rate of rise and the period of the sine wave, as follows:

Consider Fig. 201 which shows the sine wave which would appear at the output of the limiter circuit if no limiting occurred. Superimposed on this is the limited output waveform of the limiter. The limits are chosen, for convenience, symmetrically above and below the zero axis. The peak to peak amplitude of the sine wave is A_s; that of the limited rectangular wave is A_r. The period of the sine wave is t_s and the time of rise of the rectangular wave is t_r. It is required to find the time of rise of the rectangular wave in terms of the period of the sine wave and of the relative amplitudes A_s and A_r. The amplitude of the sinewave (and of the leading edge superimposed on it) varies as the sine of the angle $2\pi t/t_s$. Hence

FIG. 201.—Illustrating time of rise of limited wave.

$$\frac{A_r}{2} = \frac{A_s}{2} \sin \frac{2\pi t_r/2}{t_s} \tag{426}$$

and this leads, by rearrangement, to

$$\frac{t_r}{t_s} = \frac{\sin^{-1} A_r/A_s}{\pi} \tag{427}$$

When, as is usually the case, the time of rise is very small compared with the period of the wave, the sine may be replaced by its angle, and we obtain

$$\frac{t_r}{t_s} = \frac{A_r/A_s}{\pi} \tag{428}$$

In the circuit shown A_r is the difference between cutoff grid voltage and zero grid voltage (grid-current level). Assume a typical value, $A_r = 15$ volts, and suppose that it is required

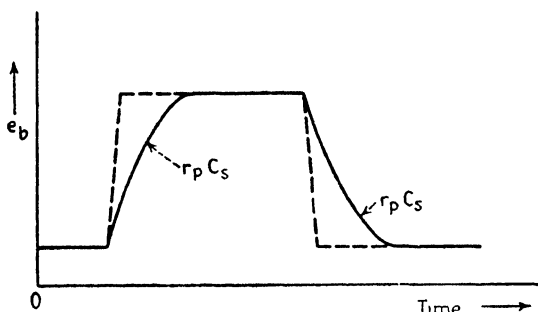



FIG. 202.—Effect of shunt capacitance on times of rise and fall.

to produce a leading edge whose time of rise is $t_r = 0.1$ μsec. from a sine wave whose period is $t_s = 0.001$ sec (frequency 1,000 cps). Then by Eq. (428) A_s must be 48,000 volts. This is much too high a voltage to be applied directly to the grid of the limiter since it would exceed the insulation and grid-current limits of the tube. It is necessary, therefore, to perform the limiting action in several successive stages. In the case cited, if the primary sine-wave source has a peak-to-peak output of 1 volt, the succeeding limiter amplifier stages must have a total gain of 48,000 (say three stages, each of gain = 36) up the grid of the final limiter stage. The design of each stage is governed by Eqs. (427) or (428).

Limiter amplification may also be applied to nonsinusoidal waveforms for the purpose of sharpening their leading and trailing edges. If the waveform is trapezoidal (leading edges of constant slope), the degree of sharpening in each stage is

directly proportional to the gain of the stage. The sharpening process cannot be carried on indefinitely, however, because the shunt capacitances at the grid and the plate of the limiter stage will eventually limit the rate of rise. By minimizing these capacitances and by employing low values of associated series or shunt resistance (including low internal plate resistance), it is not difficult to produce leading edges whose time of rise is 0.1 μsec or less. In this respect limiter stages are governed by the same considerations as multivibrators (see Fig. 202).

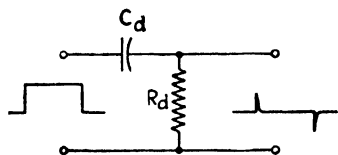



FIG. 203.—Differentiator circuit.

126. b. The Differentiator Circuit.—When a rectangular wave is available (for example, that produced by a multivibrator or by the action of limiter stages on a sinewave), the differentiator circuit (Fig. 203) may be employed to produce pulses from the leading and trailing edges of the wave. The basic circuit consists of a resistance and capacitance in series. The rectangular wave is applied across the circuit and pulses appear across the resistance. Usually the circuit is connected to the grid of an amplifier, and the combination is called a "differentiator" amplifier or "sharpener" amplifier. The differentiator circuit takes its name from the fact that the circuit responds with maximum amplitude to sudden changes in the applied voltage, but ceases to respond when the applied voltage assumes a constant value (that is, at the flat top of the applied rectangular wave). The circuit thus effectively takes the derivative of the applied voltage waveform.

FIG. 204.—Effect of source resistance and shunt capacitance on differentiator circuit.

The operation of the circuit depends not only on the basic resistance R_d and capacitance C_d, but also on the inevitable shunt capacitance C_s, and the internal resistance of the driving circuit R_s (Fig. 204). The driving circuit applies a rectangular wave of voltage to the circuit. The leading edge of the wave charges the shunt capacitance C_s along the exponential curve described by the time constant R_sC_s. Since C_s is usually small and R_s may be made small, the rate of rise is very rapid, usually a fraction of a microsecond. The circuit cannot react

faster than the voltage driving it, of course, but it can react substantially as fast.

The amplitude of the voltage appearing across C_s is the height of the pulse produced by the circuit. This amplitude is less than the applied voltage, since the capacitors C_d and C_s are charged in series and act as a capacitance voltage divider. On the assumption that R_d is large compared with R_s, the ratio of input to output voltage is

$$\frac{e_i}{e_o} = \frac{C_d + C_s}{C_d} \tag{429}$$

If C_d is substantially larger than C_s, nearly the full amplitude of the applied leading edge appears in the pulse output.

When the applied rectangular wave reaches maximum amplitude and remains constant at this level, the capacitor C_d discharges through R_d and the voltage across it decays, as shown in the figure. The time constant of the discharge is $C_d(R_d + R_s)$. The shunt capacitance C_s also discharges, producing a current in R_d in the opposite direction, decaying exponentially with time constant R_dC_s. Since C_d and R_d are generally larger, respectively, than C_s and R_s, the voltage across R_d is governed by the time constant C_dR_d. After a time $4R_dC_d$ the discharge is 98 per cent completed and the circuit soon thereafter remains quiescent until the trailing edge of the applied wave appears. The circuit then reacts as before in the opposite polarity.

By employing small values of C_d and R_d, pulses of the order of 1 μsec can be obtained, but if the value of C_d approaches that of C_s, the amplitude of the response falls off in accordance with Eq. (429). A small amplitude is often tolerated, however, since it can be increased in succeeding amplifier stages. Moreover, the differentiated pulse waveform may be passed through limiter amplifiers to obtain a more nearly rectangular shape. If desired, the negative pulses may be removed by passage through an amplifier biased to cutoff.

It must be noted that the pulse produced by a differentiator circuit is not rectangular and cannot be made precisely so even by passage through limiter stages. A closer approach to rectangular shape can be achieved if resonance is introduced by the insertion of a small inductance ($L = \frac{1}{2}C_sR_d^2$) in series with the resistor R_d. The circuit then produces a pulse somewhat wider

at the top, and with a steeper trailer edge, than that produced by the simple RC circuit.

127. c. Synchronized Multivibrators and Blocking Oscillators. The leading or trailing edge of an extended waveform may be employing to trigger a pulse generating circuit,' such as the unsymmetrical multivibrator or blocking oscillator. The advantage of this procedure is that the pulse shape is independent of the shape of the control waveform, depending simply on the properties of the pulse-forming circuit previously discussed (Secs. 123 and 124).

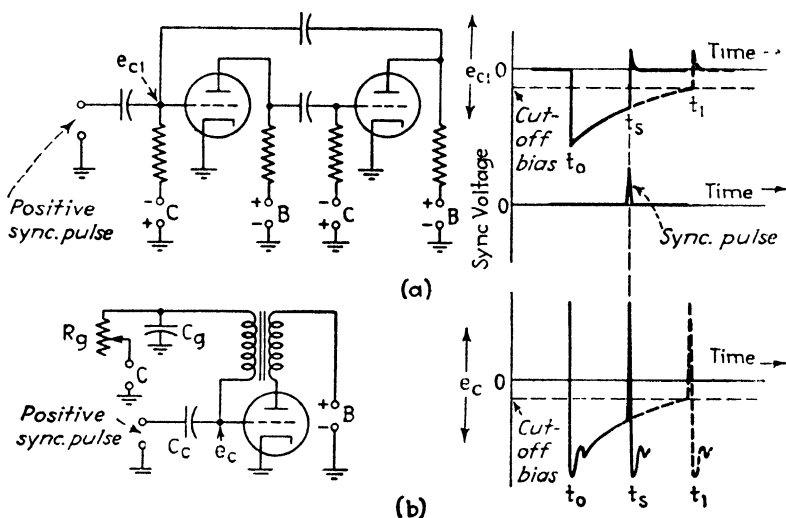



FIG. 205.—Multivibrator (*a*) and blocking oscillator (*b*) synchronized by externally-applied pulses.

Figure 205 shows typical synchronized pulse generators and their waveforms. The essential factor in each circuit is a negative bias voltage on the synchronized grid. If the magnitude of this bias voltage is large (that is, tube grid biased below cutoff), the pulse generator remains quiescent in the absence of the synchronizing voltage. When the synchronizing voltage raises the grid above cutoff, the pulse is formed. A smaller value of bias, or no bias at all, suffices if the synchronizing voltage is employed merely to correct the natural frequency of the pulse generator. This action is used in frequency multiplying and dividing circuits (Sec. 133).

128. d. Pulse-forming Transmission Lines.—A circuit widely used to form pulses, particularly at high power level, is the combination of an electronic switch and an artificial transmission line or low-pass filter. The electronic switch is a means of connecting a load suddenly to a charged transmission line; the switch may be a spark gap, a gas-discharge tube controlled by a rectangular wave or pulse, or a high-vacuum tube so controlled. A typical line-controlled pulse generator is shown in Fig. 206. The artificial transmission line is composed of a large enough number of sections (usually four or more) to approximate a uniform transmission line (cf. Sec. 64, Chap. III).

FIG. 206.—Use of an artificial transmission line to form pulses.

The power supply charges the transmission line to the voltage amplitude E prior to the closing of the electronic switch. When the switch closes, the line discharges and the voltage across it divides equally between the line itself and the terminating resistance R_0, since R_0 is chosen equal to $\sqrt{L/C}$, the characteristic impedance of the line. Thus a voltage of $E/2$ suddenly appears across R_0, forming the leading edge of the pulse. The other half of the voltage travels as a wave down the transmission line, partially discharging the line as it propagates. When the wave reaches the end of the transmission line it is totally reflected (since the line is effectively open circuited at this point) without change of polarity. Returning along the line, the wave completes the discharge of the line. When the wave arrives back at R_0 the voltage across R_0 suddenly collapses as the discharge is completed, and the trailing edge of the pulse is thereby formed. The circuit thereafter may be recharged and the cycle repeated.

Essentially the line-controlled pulse generator is an echo device. The length of the pulse generated depends on the time it takes the voltage wave to traverse the transmission line and return. The voltage wave is delayed $\sqrt{LC}$ μsec in passing each section of the line. If the line contains N sections, the width of the pulse generated is

$$d = 2N\sqrt{LC} \qquad \mu\text{sec} \tag{430}$$

where L is the series inductance of each section in microhenries

and C is the shunt capacitance of each section in microfarads. The circuit is well adapted to the formation of the short pulses required for modulating purposes, but not for the formation of longer rectangular waves because of the difficulty of designing a uniform artificial line with long delay.

The practical difficulty in any event is the large number of sections required to approximate a uniform line. If a small number of sections is used, the approximation is a poor one and the flat top of the pulse displays irregularities corresponding to the separate resonances of the individual filter sections. These

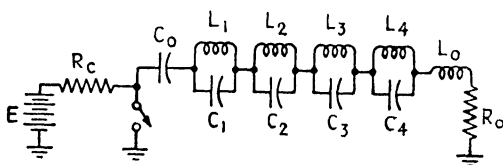



FIG. 207.—Parallel transformation of transmission line (Guillemin line).

irregularities may be removed by passing the pulse through a limiter stage, but in high-power circuits limiting is an undesirable complication.

A refinement of the low-pass filter design has been devised by Guillemin to secure a more regular flat top with a given number of filter sections. An empirical study established the fact that the following trigonometric series represents a rectangular wave with a top of maximum flatness, when the line is limited to five sections.

$$W(t) = 1.258 \sin \omega t + 0.373 \sin 3\omega t + 0.174 \sin 5\omega t + 0.083 \sin 7\omega t + 0.028 \sin 9\omega t \quad (431)$$

Each term in the series can be represented by a series LC circuit, the L and C values being chosen such that $\sqrt{C/L}$ equals the coefficient of the term and $1/(\sqrt{LC})$ the angular frequency of that term [cf. Eq. (44), Chap. II].

The transmission line thus formed is transformed into the equivalent series network of shunt LC circuits. The latter form of line is simpler to realize with practical circuit elements and permits the use of less bulky and expensive capacitors, since the applied voltage is divided nearly equally among them. A typical Guillemin-line generator designed on this basis is shown in Fig. 207.

The artificial transmission can be used in a variety of circuits other than the basic one just considered. In the cathode-line circuit (Fig. 208*a*), the synchronizing voltage drives the grid positive and a corresponding positive wave travels down the transmission line and returns to the cathode with double amplitude, which is sufficient to cut off the plate current. The plate

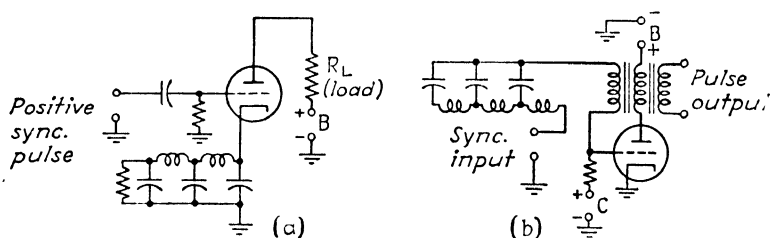



Fig. 208.—Pulse-forming transmission lines in cathode circuit (*a*) and blocking oscillator (*b*).

current thus starts and stops at an interval, governed by the length of the transmission line, given in Eq. (430). The line-controlled blocking oscillator (Fig. 208*b*) operates on a similar basis. The synchronizing waveform charges the line and at the same time brings the grid of the oscillator above cutoff. The oscillator reacts and would continue to do so, producing a long pulse. But the wave reflected from the transmission line cuts off the grid before the longer pulse is completed and a sharp trailing edge is thereby produced. Thereafter the grid recovers along the conventional exponential curve to a value below the cutoff level and there remains until the next synchronizing impulse arrives.

Fig. 209.—Single-capacitor pulse generator, using high-vacuum switch tube.

A pulse circuit widely used for modulation in early designs is the simple capacitor discharge circuit shown in Fig. 209. This is closely related to the line-controlled circuit, the artificial transmission line being replaced by a capacitor. The capacitor is charged from the d-c supply while the electronic switch is open (between pulses). When the electronic switch is energized, the capacitor discharges (partially at least) through the series circuit composed of the load (across which the

pulse appears) and the switch. The inductance in series with the power supply prevents current flow from the power supply while the electronic switch is closed. When the switch reopens, at the conclusion of the pulse, the capacitor is charged at a slow rate by the power supply.

Several variations in the method of charging the artificial transmission line (or single capacitor, if this is used) are possible. The primary power source may be of the a-c type, with the power frequency chosen to equal the desired pulse rate. The electronic switch is operated synchronously with the power frequency and in such phase that the line or capacitor is discharged at the peak

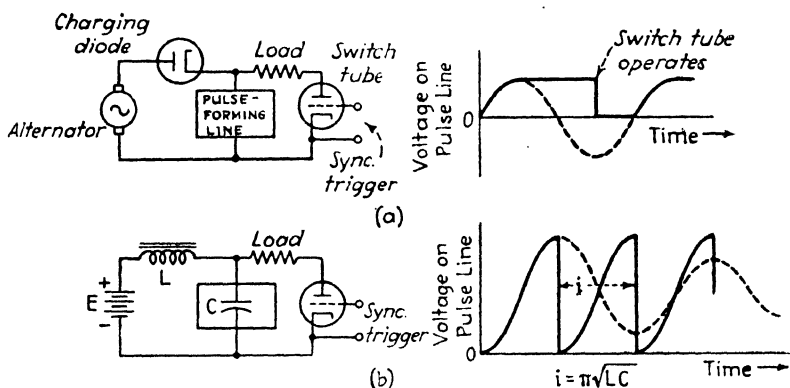



FIG. 210.—Use of synchronous alternator (*a*) and series resonance (*b*) in pulse generators.

of the positive charging cycle (Fig. 210*a*). A diode is necessary to prevent discharge during the negative half of the primary power cycle.

Another variation (Fig. 210*b*) employs a d-c source, with the inductance and capacitance chosen to resonate at a frequency equal to one-half the desired pulse rate. When the electronic switch is opened, at the conclusion of the pulse, the direct voltage of the power supply is suddenly applied to the series LC circuit and oscillations are set up. The electronic switch is closed at the end of the first half-cycle of this resonant oscillation, forming the pulse when the voltage across the capacitor has its maximum value, which is about twice that of the power supply voltage. This is a simple means of obtaining a 2:1 increase in the pulse voltage relative to the power supply voltage.

In all the line-controlled circuits, complications ensue due to

the presence of small stray inductances, which tend to induce the oscillations at the conclusion of the pulse. These oscillations may be damped out by the use of auxiliary diodes and damping resistors. These damping devices are discussed in the sections on specific modulator circuits (Chap. IX).

129. e. Saturable Core Generators.[1]—A direct means of forming pulses from a sinusoidal timing source is the use of an iron-core reactor whose core saturates at a low level of magnetization. The magnetization curve of such a core is shown in Fig. 211. Whenever the current in the reactance exceeds the limits cor-

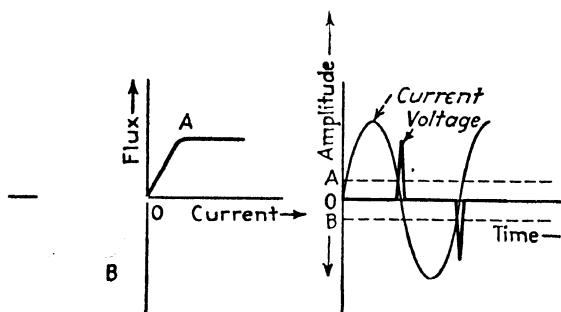



FIG. 211.—Saturable-core reactor as a pulse generator.

responding to points A and B, the magnetic field fails to increase and the voltage induced across the coil (which depends on rate of change of flux in the core) drops to zero. When a sinusoidal current of large magnitude is applied to such a reactor, voltage appears across the coil only during those portions of the wave corresponding to the limits A and B, as shown. The width of the pulse thus generated, relative to the period of the applied wave, depends on the maximum amplitude of the applied current, relative to the current at which saturation occurs.

The saturable core reactor may, of course, be excited by a non-sinusoidal wave, and it will produce pulses shorter than the applied wave when the amplitude of the excitation exceeds the saturation current limits shown in the figure.

The device may be considered as a nonlinear impedance, of nominal value when the current lies within the saturation limits and of much lower value when the current lies outside these limits.

[1] PETERSON, E., Coil pulsers for radar, *Bell Sys. Tech. J.*, **25**, 603 (October, 1946).

Acting on this principle, the reactor may be used as a switching device to discharge the pulse-controlled circuits previously discussed. Figure 212 shows the basic circuit. An auxiliary current-control tube (a beam-power pentode tube) drives the reactor from one saturation limit to the other, varying its impedance. Between the saturation limits, the pulse-forming capacitor is discharged through the reactor.

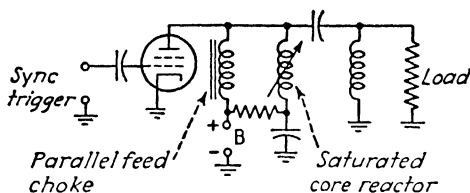



FIG. 212.—Saturable reactor pulse-generator circuit.

130. Synchronized Production of Extended Waveforms.—The synchronized production of extended waves has already been discussed in connection with the deformation of sine waves and the synchronized multivibrator. In the first case, the synchronism is provided inherently by the sine wave; in the second, it depends on the response of the circuit to a positive impulse that raises the grid of the nonconducting tube above the cutoff level. These methods are suitable for the production of rectangular waves; in addition, the multivibrator produces exponential waves at its grids. Two other synchronized extended waveforms are of interest: (1) rectangular waves produced by the direct-coupled multivibrator, and (2) sawtooth waves.

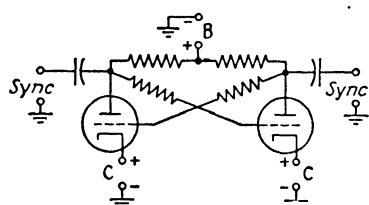



FIG. 213.—Direct-coupled multivibrator (Eccles-Jordan circuit).

The direct-coupled multivibrator (developed by Eccles-Jordan and often referred to by his name) is illustrated in Fig. 213. Each grid is connected conductively, through a resistance, to the plate of the opposite tube. Since there are no coupling capacitors, the grid of the cutoff tube does not recover exponentially, as in the conventional multivibrator. Rather the cutoff grid is maintained in the cutoff condition by the depressed plate potential of the opposite tube. The circuit thus remains in stable equilibrium indefinitely, until a positive synchronizing

voltage is applied to the cutoff grid. When this grid is brought above the cutoff level, the plate current of the same tube rapidly rises, forcing the opposite tube into the cutoff condition, where it remains until a synchronizing impulse is applied to its grid. The circuit thus produces one full cycle of operation for every two applied synchronizing impulses, that is, it inherently divides the synchronizing frequency by a factor of 2.

The waveforms generated at plate and grid are rectangular, since there is no exponential discharge. The leading and trailing edges of the waves are governed by the shunt capacitance and associated series resistance, as in the conventional multivibrator circuit. In practice, the resistors between grid and plate are often by-passed with capacitors to minimize the effect of the series resistance on the rate of rise. The rectangular wave is symmetrical in time when the synchronizing pulses are evenly spaced. The circuit will follow irregularly spaced pulses that must be employed if rectangular waves of unequal width are required.

Among the many variations of the synchronized multivibrator is the one shown in Fig. 214, known as the "one-shot" multivibrator. One of the grid-plate connections is conductively coupled; the other is capacitively coupled. The synchronizing impulse is applied to the conductively coupled grid. The circuit responds when this grid is brought above the cutoff level, driving the opposite tube to the nonconducting state. This tube recovers, by virtue of the discharge of the capacitor connected to its grid, reaches the conducting level, and drives the first (synchronized) tube to cutoff. The latter tube remains in the cutoff condition indefinitely until another synchronizing impulse appears. The circuit thus reacts once to each synchronizing impulse, and will follow the synchronizing impulses in any sequence, provided that the impulses are separated by longer intervals than the recovery time of the circuit.

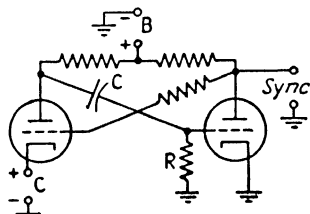



FIG. 214.—One-shot multivibrator or trigger circuit.

131. Synchronized Production of Sawtooth Waves.—The sawtooth wave finds its principal use in radar in the deflection of the cathode-ray beam in the indicator. The deflection along the

range coordinate starts synchronously with the transmitted pulse and continues, at as nearly constant a rate as possible, during a specified portion of the pulse interval (that portion during which echoes are to be received). At the end of the deflection period, the indicator beam is returned rapidly to its starting point and remains there until the next transmitted pulse. The general form of such a deflection wave is shown in Fig. 215*a*.

A sawtooth wave of voltage is required when electric deflection of the cathode-ray beam is used; a current wave is necessary to create magnetic deflection. Generally speaking, the two cases require separate treatment, since a sawtooth wave of voltage does not produce a sawtooth wave of current except in a resistive circuit. The deflection coil of a magnetic deflection system is primarily inductive.

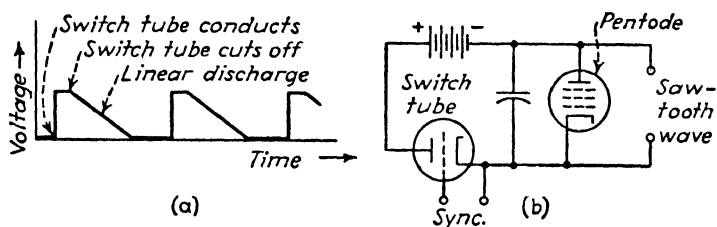



FIG. 215.—Sawtooth wave (*a*) and constant-current discharge sawtooth generator (*b*).

Consider first the production of sawtooth waves of voltage. The primary requirements are precision of timing and linearity of the deflection. Precise timing is obtained by the use of a sharp leading edge in the synchronizing impulse. The linearly increasing voltage is obtained by charging a capacitance with a steady current.

The basic circuit used to obtain linear sawtooth waves of voltage is shown in Fig. 215*b*. It consists of a capacitor that is charged or discharged at constant current through a pentode tube. The pentode has the property of passing a substantially constant current while the voltage applied to it varies within wide limits. In the circuit shown, the capacitor is charged by the application of the battery voltage through the electronic switch, which operates under the control of an external synchronizing voltage. When the switch is deenergized, the capacitor discharges through the pentode, and the voltage across it decays linearly, as shown in the accompanying waveform.

Since the discharge is not completely linear at the base of the curve, the corresponding portion of the deflection may be deleted or disregarded. After the discharge is completed, the capacitor voltage remains at zero until the electronic switch is again closed and the cycle repeats. The corresponding synchronizing waveform is shown in the figure. By choosing the relative length of the on and off portions of the curve it is possible to vary the shape of the sawtooth wave.

For magnetic deflection, the sawtooth wave of current is produced by applying a pulse waveform to the inductive deflection yoke.

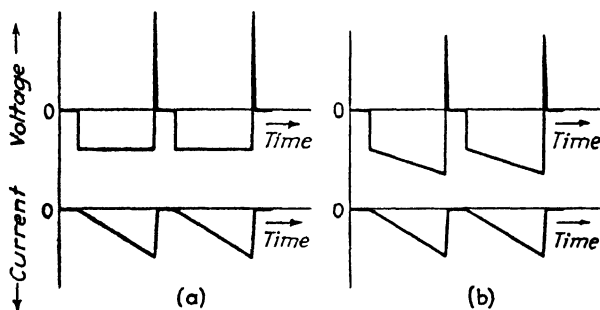



FIG. 216.—Voltage waves in an inductive circuit (a) and a resistive-inductive circuit (b) required to produce sawtooth waves of current.

Since a suddenly applied voltage across an inductance gives rise to a steadily increasing or decreasing current [see Eq. (33), Chap. II], the leading and trailing edges of the pulse give rise to linear currents, as shown in Fig. 216a. In practice, the deflection coils display a combination of resistance and inductance, which modifies the required voltage waveform as shown in Fig. 216b. The production of the necessary synchronizing pulses involves the circuits previously discussed (Secs. 123 to 129). The blocking oscillator, followed by a high-current pulse amplifier tube, is often used.

132. Modification of Waveform Shape.—When a pulse has been generated, either directly or in synchronized circuits, it is often necessary to perform one or more modifications of its shape to meet particular needs. Two basic circuits functions used for this purpose have already been described: differentiation and limiting. Two other modifications are of interest: integration of a series of pulses, and removal of unwanted polarity.

Integration is carried out by feeding a succession of pulses to a

capacitor, through a series resistance, as shown in Fig. 217. An essential property of the circuit is its *RC* time constant, which is long compared with the duration of each pulse. If the pulses are symmetrical, each pulse is merely broadened and lengthened. But if a rectifier is interposed in series with the capacitor, preventing discharge at the conclusion of each pulse, the voltage across the capacitor builds up in steps until it attains a value

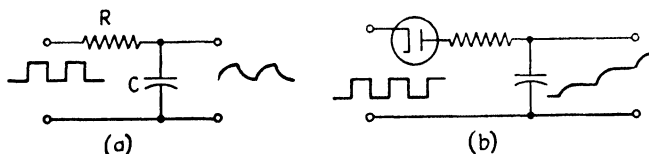



FIG. 217.—Integrator circuit (*a*) and diode accumulator circuit (*b*).

sufficient to prevent passage of the current through the rectifier. This action is identical in principle to the peak detector employed in radio receivers. The two cases are illustrated in the figure.

Often it is desired to produce pulses only at the leading edge of the synchronizing wave, although the trailing edges must inevitably accompany the leading ones. In symmetrical pulse-forming circuits, like the differentiator or saturable core reactor, the trailing edges produce pulses of opposite polarity. The usual method of removing pulses of unwanted polarity is to pass the pulse sequence, in proper polarity, through an amplifier biased to cutoff. Another equally effective circuit, which will remove pulses of either polarity, is the double-diode limiter, shown in Fig. 218. The shunt diode passes pulses of one polarity to ground, while the series diode passes the pulses of opposite polarity to the load circuit. An important property of this circuit is the fact that the load circuit is completely isolated, so far as the reverse polarity is concerned, from the driving circuit. The pulses of unwanted polarity need not be removed completely. In common with other limiter circuits, the double-diode circuit may be set to limit at any level by the inclusion of biasing voltages, as shown in the figure.

FIG. 218.—Biased diode circuit for eliminating negative or positive pulses (shown connected to remove negative pulses).

Figure 219 shows a typical sequence of pulse-shaping operations involving differentiation, limiting, integration, and pulse removal with the corresponding circuits and the associated time constants. Many variations are, of course, possible in the application of these circuits.

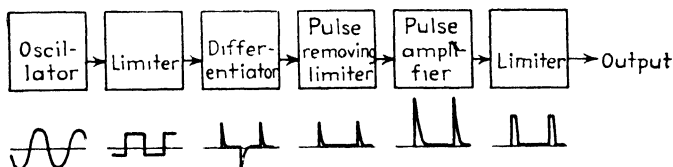



FIG. 219.—Typical sequence of pulse-shaping operations.

133. Pulse-rate Multiplication and Division.—Multiplication and division of repetition frequency are operations that are occasionally required in pulse circuits. Frequency division is needed, for example, to reduce the sine-wave frequency of a piezoelectric oscillator to pulse frequency. Frequency multiplication is used to produce a series of pulses, for calibration purposes, from a lower basic pulse frequency.

Among the frequency divider circuits are the synchronized multivibrator, and the counter circuit. Frequency multiplica-

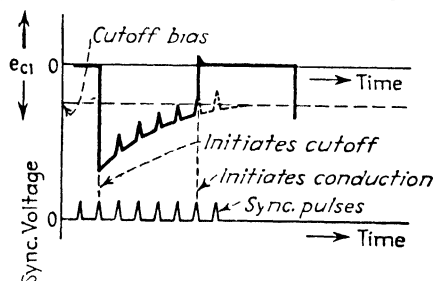



FIG. 220.—Grid waveform of a synchronized symmetrical multivibrator used as a frequency divider.

tion is commonly carried out, in nonsinusoidal circuits, by the use of the synchronized multivibrator and the shock-excited oscillator.

The synchronized multivibrator, used as a frequency divider, is illustrated in Fig. 220. The action of this circuit involves the application of a number of input synchronizing pulses across the control grid of the cutoff tube in a biased multivibrator. When the sync voltage, superimposed on the recovery curve,

causes the grid potential to rise above the cutoff level, the multivibrator oscillates and after one cycle returns to the initial state, where it is ready to receive the integrated effect of another group of sync pulses. A number of pulses, say n, may be applied to the input for each output cycle, corresponding to frequency division of n. Operation is most stable when the division is limited to a low factor, say 5 or less. The successive values of grid voltage are then separated by larger steps, relative to voltage variations in the input circuit arising from other causes. In a sequence of such divider circuits, it is desirable

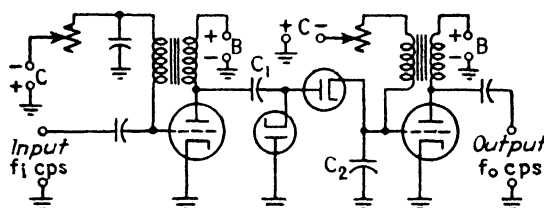



FIG. 221.—Blocking oscillators and counter circuits as frequency dividers.

to operate each divider at an odd factor (3, 5, or 7 are typical values).

The counter circuit, illustrated in Fig. 221, operates on a similar principle but employs a synchronized blocking oscillator in place of the multivibrator, and the integrator circuit is of the diode type. The first blocking oscillator operates at a frequency of f_i pulses per sec. Each pulse has an amplitude equal to about 1.5 times the voltage of the B supply and charges capacitors C_1 and C_2 in series. Hence, capacitor C_2 receives an increment of voltage, at each input pulse, equal to

$$E_{C2} = \frac{1.5E_{bb}C_1}{C_1 + C_2} \tag{432}$$

Since discharge is prevented by the diode, after n input pulses the voltage applied to synchronize the second blocking oscillator is $n\,\Delta E_{C2}$. When this voltage exceeds the cutoff limit, the second blocking oscillator responds, and its frequency of oscillation is

$$f_0 = \frac{f_i}{E_s/\Delta E_{C2}} = \frac{f_i}{n} \tag{433}$$

where E_s is the voltage required to trigger the second blocking oscillator.

When the second blocking oscillator reacts, it completely discharges capacitor C_2 (during the grid-current surge), and C_2 is thereafter ready to receive further increments of charge from the preceding circuits. One of the advantages of this circuit is the fact that all critical voltage levels are proportional to the B supply voltage, and hence the circuit tends to count the same number of pulses when the B supply voltage varies. Properly designed counter circuits remain stable when the plate supply is varied from 200 to 300 volts, but this is true only if the number of pulses counted is small, say 7 or less, for the reasons already stated in connection with the multivibrator divider circuit.

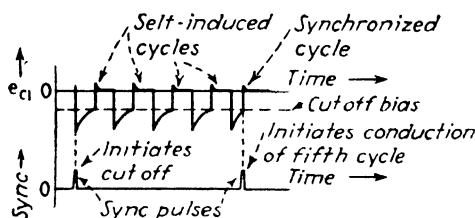



FIG. 222.—Waveform of a synchronized symmetrical multivibrator as a frequency multiplier.

The multivibrator when used as a frequency multiplier (Fig. 222) is essentially a self-excited oscillator, which continues to operate in the absence of the synchronizing voltage. The synchronizing signal is applied at a lower frequency, one input pulse for every n cycles of the oscillator, and in such phase that it anticipates the formation of the nth cycle. The nth cycle is thus forced to occur in synchronism with the input sync pulse. By proper adjustment of the free-running oscillation frequency, it is possible to arrive at a frequency that is just a multiple of the input frequency. Thereafter the input signal corrects the free-running frequency and maintains synchronism. This circuit is somewhat critical of adjustment, and the period of the nth cycle tends to be somewhat shorter than the intervening periods; and therefore the circuit is not used in precision calibration circuits, where each period must be accurately equal to every other.

The shock-excited oscillator (Fig. 223) is a discontinuous frequency multiplier that forms a succession of accurately timed sine waves for each application of the synchronizing voltage. The oscillator is an LC shunt-tuned circuit loaded with

resistance. The current in such a circuit when excited by a unit pulse of voltage is

$$i = \frac{E}{\omega L} \epsilon^{-Rt/(2L)} \sin \omega t \tag{434}$$

where

$$\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \tag{435}$$

The waveform corresponding to this equation is plotted in the figure. Oscillation occurs when the series resistance R has a value less than the critical value $2\sqrt{L/C}$. The time constant of the envelope of the oscillations is $2L/R$. The damped sine

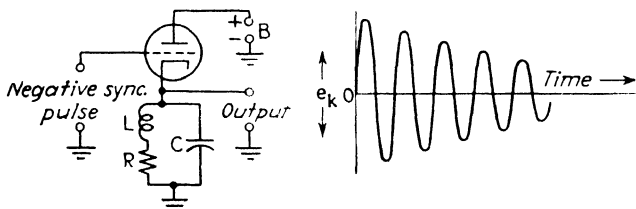



FIG. 223. Shock-excited oscillator.

wave is converted into pulses, for calibration purposes, by the deforming and differentiating circuits previously described.

Frequency multiplication may also be performed on non-damped sinusoidal sources, using conventional frequency multiplying (nonlinear doubler or tripler) amplifiers. Little use of this method is made in pulse circuits, however, since in most practical cases the need is to reduce the frequency of continuous sinusoidal sources, rather than to increase it.

134. Amplification of Pulse Waveforms. Linear Video Amplification.—Linear video amplification is a circuit function intended to increase the voltage, current, or power level of a pulse or extended waveform without materially changing its shape. The term "video" comes from television practice; as usually defined, it denotes amplification prior to modulation and subsequent to demodulation. The term "linear" denotes direct proportionality between input and output waveforms, that is, maintenance of pulse shape. The limiter amplifiers and differentiator amplifiers previously discussed are nonlinear amplifiers.

The basic linear video amplifier circuit is that shown in Fig. 224. Since in linear amplification the amplifier tube operates

over the linear portion of its dynamic characteristic, it is permissible to discuss the circuit action in terms of variational parameters of the tube, that is, amplification factor μ, grid-plate transconductance g_m, and internal plate resistance r_p. Over the v-f (video-frequency) range (zero to, say, 10 megacycles) these quantities are independent of frequency. Hence

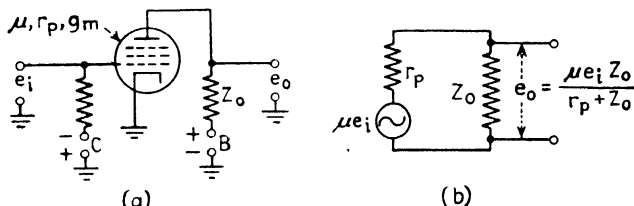



FIG. 224.—Actual and equivalent circuits of the linear video amplifier.

the frequency response of a video amplifier is determined solely by the magnitude and phase of the load impedance Z_0, as functions of frequency. The voltage gain of a video amplifier is defined as the ratio of the output voltage across the load impedance to the input voltage across the grid and cathode. The equivalent circuit in Fig. 224 shows that the gain so defined is

$$G = \frac{e_0}{e_i} = \frac{\mu Z_0}{r_p + Z_0} \tag{436}$$

In most practical cases, the internal plate resistance r_p is large compared with the resistive component of Z_0, and Z_0 is primarily resistive over the v-f range. Then the expression for the gain reduces to

$$G = \mu \frac{Z_0}{r_p} = g_m Z_0 \tag{437}$$

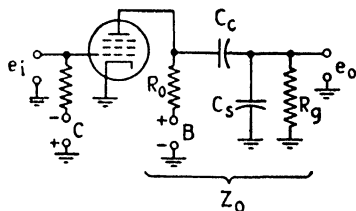



FIG. 225.—Component elements of the load impedance Z_0.

To investigate the operation of a video amplifier, it is convenient to examine the quantity Z_0 over the frequency band that comprises the essential frequencies of the pulse spectrum. This is the steady-state method of analysis discussed at length in Chap. II.

The simplest form of capacitance-coupled video amplifier is the so-called "uncompensated type" shown in Fig. 225. The load impedance Z_0 consists of two resistances, R_0 and R_g, and

two capacitances C_c and C_s connected as shown in the figure. The last named is the total shunt capacitance, which is of interest only at the h-f end of the band. At high frequencies the several shunt capacitances present can be lumped in one value (C_s) since they are separated by a negligible reactance (that of C_c).

It is convenient to discuss the magnitude and phase of this impedance in two frequency regions, the high and the low. At

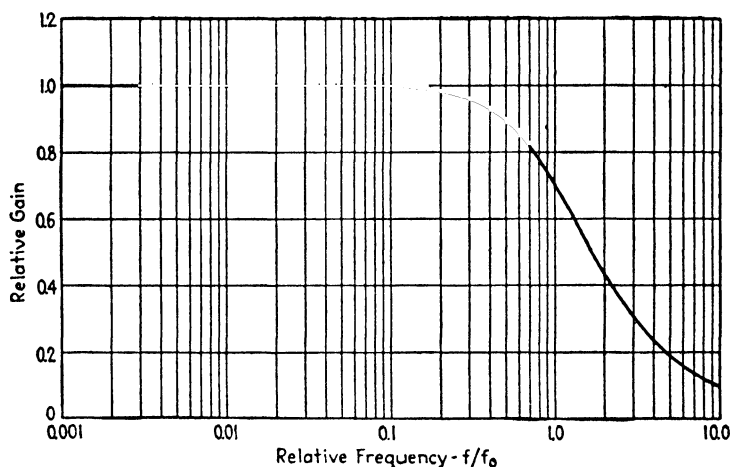



FIG. 226.—High-frequency amplitude response of the uncompensated video amplifier.

the high frequencies the magnitude of the impedance is given by

$$Z_0 = \frac{R_t}{\sqrt{1 + f^2/f_0^2}} \tag{438}$$

where R_t is the resultant resistance formed by R_0 and R_g in shunt, and f_0 is

$$f_0 = \frac{1}{2\pi R_t C_s} \tag{439}$$

A plot of Eq. (438) is given in Fig. 226. It will be seen that f_0 is an index of the upper frequency limit of the band passed by the amplifier. Actually, it is the frequency at which the gain of the amplifier drops to 70.7 per cent of its value at middle frequencies. To increase the upper frequency limit of the amplifier, it is necessary to employ small values of R_t or C_s or both. The lower limit of C_s is reached when all shunt capacitance is removed

except the inevitable interelectrode capacitances and those of the wiring and circuit elements between tubes. The lower limit of R_t is $1/g_m$, since at this value the m-f (middle-frequency) gain of the stage reduces to unity. Practical values in circuits employing receiving tubes are R_t, 2,000 ohms and C_s, 20 $\mu\mu$f.

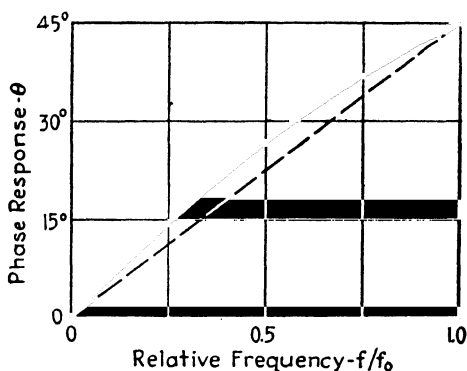



FIG. 227.—High-frequency phase response of the uncompensated video amplifier.

The corresponding value of f_0 is, by Eq. (439), about 4 megacycles. This frequency corresponds to the first zero in the spectrum of a rectangular pulse whose width is 0.25 μsec. If pulses shorter than this are to be amplified, or if the fidelity required is such that the band passed must include the second or third zero of the pulse spectrum, a higher value of f_0 must be achieved. C_s must be minimized and the tube used must display the highest possible value of g_m, thus permitting the use of low values of R_t without excessive loss of gain.

FIG. 228.—Elements of load impedance at low frequencies.

The phase of Z_0 at high frequencies is given by

$$\theta = \tan^{-1} - \frac{f}{f_0} \qquad (440)$$

We recall from Chap. II that the phase characteristic of a distortionless amplifier must be closely linear over the pulse spectrum. Figure 227, a plot of Eq. (440), shows the degree by which the uncompensated amplifier falls short of the ideal.

The l-f (low-frequency) response of this circuit is governed by the equivalent circuit shown in Fig. 228. C_s now has a negligible

effect, but C_c cannot be neglected. The magnitude of Z_0 at low frequencies is given by

$$Z_0 = \frac{R_0(f/f_c)}{\sqrt{1 + (f/f_c)^2}} \tag{441}$$

where f_c is a reference frequency denoting, approximately, the ower frequency limit of the amplifier.

$$f_c = \frac{1}{2\pi R_g C_c} \tag{442}$$

To maintain the l-f response of the amplifier, large values of R_g or C_c, or both, should be used. The limit in this respect is placed

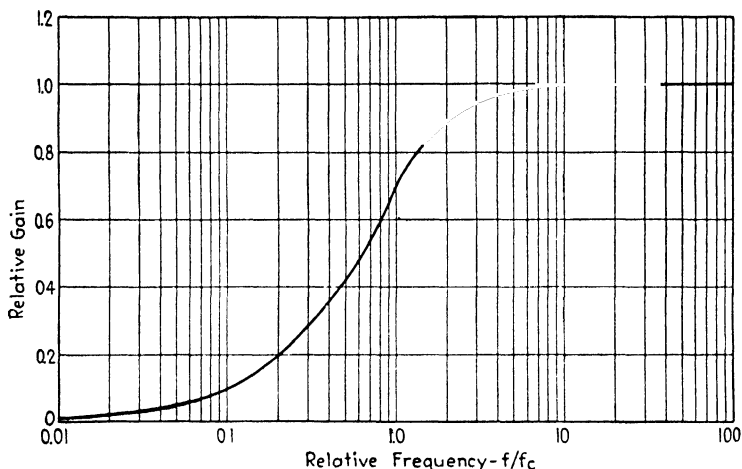



FIG. 229.—Low-frequency amplitude response of the uncompensated video amplifier.

by the tendency of the amplifier to oscillate in relaxation fashion by coupling through the power supply impedance, if too large a value of C_c is used. A large coupling capacitor also contributes to the shunt capacitance and hence mitigates against good h-f response. A plot of Eq. (441) is shown in Fig. 229.

The phase angle of Z_0 at low frequencies is given by

$$\theta = \cot^{-1} \frac{f}{f_c} \tag{443}$$

The departure from linearity is marked (Fig. 230), but the effect on the fidelity of reproduction is not marked when the pulse width is short compared with the pulse interval. When extended

l-f rectangular or sawtooth waves are amplified, however, a nonlinear l-f phase angle has a pronounced effect.

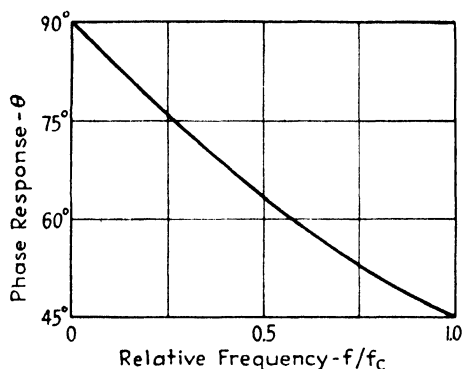



FIG. 230. —Low-frequency phase response of the uncompensated video amplifier.

135. Compensation of Video Amplifiers. High-frequency Compensation.—The h-f limit of a video amplifier stage may be extended by the inclusion of inductive elements in the load impedance. A widely used circuit is the shunt-compensated connection shown in Fig. 231. A detailed analysis of this circuit has been given in Sec. 49, Chap. II, to which the reader should refer. The inductance is usually chosen as

FIG. 231. Shunt inductive high-frequency compensation.

$$L = 0.5C_sR_l^2 \tag{444}$$

The magnitude of the impedance Z_0 is then maintained equal

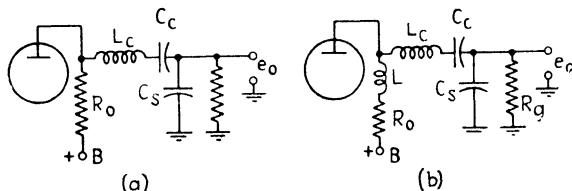



FIG. 232.—Series (a) and shunt-series (b) high-frequency compensation.

to its m-f value, up to the frequency f_0 whereas in the uncompensated amplifier the magnitude drops to 70.7 per cent of its m-f value. The cutoff above f_0 is somewhat sharper in the

compensated circuit, however, and this may be undesirable for the reasons stated in Chap. II, Sec. 53.

Other forms of inductive compensation may be used. Typical examples are shown in Fig. 232. The series and series-shunt circuits are essentially forms of filter networks. The optimum

TABLE X.—HIGH-FREQUENCY COMPENSATION METHODS

Type	R_0	L	L_c	Relative gain at f_0	Variation in time delay up to f_0
Uncompensated	$1/(2\pi f_0 C_s)$	....	....	0.707	$0.035/f_0$
Shunt	$1/(2\pi f_0 C_s)$	$0.5C_sR_t^2$	....	1.0	$0.023/f_0$
Series	$1.5/(2\pi f_0 C_s)$	....	$0.67C_sR_t^2$	1.5	$0.011/f_0$
Shunt-series	$1.8/(2\pi f_0 C_s)$	$0.12C_sR_t^2$	$0.52C_sR_t^2$	1.8	$0.015/f_0$

circuit for this purpose has been shown by Wheeler[1] to be a filter of several sections, but this circuit suffers in practice from the

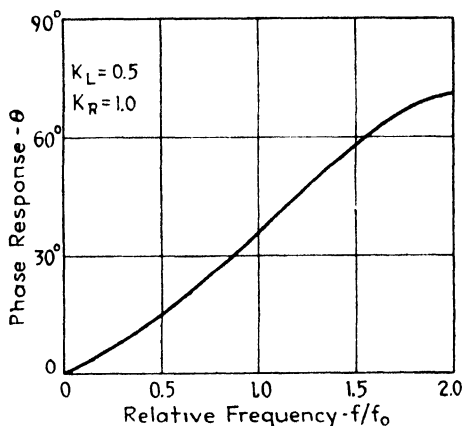



FIG. 233.—High-frequency phase response of compensated amplifier (compare Fig. 227).

critical nature of the required adjustment. The relative design factors and performance of the several forms of inductive compensation are given in Table X.

The phase response of the shunt-compensated amplifier is given by Eq. (98), Chap. II. Figure 233 shows the phase

[1] WHEELER, H. A., Wideband Amplifiers for Television, *Proc. I.R.E.*, **27**, 429 (July, 1939).

function for the case discussed immediately above ($K_L = 0.5$ and $K_R = 1.0$). Comparison with Fig. 227 shows the improvement in phase linearity resulting from the introduction of the shunt inductance. Table X gives corresponding information on the phase performance of the other compensation circuits.

136. Low-frequency Performance.—The l-f response of video amplifiers is not critical if the pulse rate is 200 pps or higher (as it usually is in radar practice) and if the pulse width is small compared with the pulse interval. For some purposes, particularly the amplification of extended l-f rectangular waves, compensation of l-f response is desirable. The insertion of a decoupling filter, shown in Fig. 234, serves to permit the use of large values of R_g and C_c without danger of relaxation oscillations. Moreover, the phase response is made more nearly linear by the insertion of the decoupling filter. The design is usually based on the following relationship:

$$C_cR_g = \frac{C_fR_0R_f}{R_0 + R_f} \qquad (445)$$

FIG. 234.—Low-frequency compensation.

An important aspect of the l-f design of a video amplifier is the overload limit. When positive pulses of sufficiently large magnitude are applied to the grid of a video amplifier stage, grid current flows and charges the coupling capacitor. At the conclusion of the pulse, the voltage thereby built up across the coupling capacitor remains and may depress the grid of the stage below the cutoff level. The stage is thereby blocked and will not respond to further signals until the voltage across C_c has discharged through the associated series resistance R_g. This blocking effect is particularly troublesome when it is necessary to amplify echoes from near-by targets, that is, those following the transmitted pulse by a very short interval.

The blocking voltage increases with the magnitude and duration of the pulse, as well as with the magnitude of the grid current. All these quantities should be minimized, where possible, to avoid building up sufficient voltage to block the stage. The use of an amplifier tube having a highly negative cutoff (remote cutoff) is also indicated. Small values of C_c

and R_g should be used to assist in rapid recovery from the blocked condition. Special circuits have been developed to reduce the blocking time to a few microseconds. For example, a separate pulse originating in the timer may be applied synchronously to deactivate the amplifier stage during the transmitted pulse.

137. Current and Power Relations in Video Amplifiers.—The video amplifier has been discussed thus far merely as a voltage-amplifying device. For certain applications, particularly driving magnetic deflection systems, current amplification is the basic function. Current amplification differs from voltage amplification only in degree. The current flowing in the plate circuit of the preceding amplifier is taken as the reference and the current in the plate circuit of the current-amplifier stage compared with it. The requirements of such a stage are evident: the tube should have a high value of g_m, high plate-current capacity, and should, of course, feed a low-resistance load circuit.

In modulation circuits, power amplification is the basic function. Since the modulated r-f generator generally has a specific input impedance value, the power generated in the power amplifier must appear across a specified resistance. This is equivalent to specifying the voltage amplification or alternatively, the current amplification. Power video amplifiers generally must be operated at high voltage and must have high plate-current capacity. Often many tubes must be used in parallel to meet the current requirement. This increases the shunt capacitance arising from interelectrode sources, but increases the effective grid-plate transconductance in the same proportion; and thus the h-f performance of a multitube power video amplifier is not substantially worse than that of a single tube.

138. Cathode-coupled Video Amplifiers. a. The Bootstrap Circuit.—An important aspect of pulse circuit design is economy in the use of power supply and vacuum tubes. Since the duty cycle of most radar pulse sequences is very small, economy is served by driving an amplifier with positive pulses, so that plate current flows only during each pulse. The average plate current, and the corresponding emission and plate power dissipation required of the tube, are then at a minimum. If the opposite polarity is employed, current flows during the pulse

intervals and the power consumption and required size of the vacuum tubes are both greatly increased. At low level these considerations are not too serious, but in high-level circuits, employed in and just prior to the modulation stage, pains must be taken to operate each of several cascaded stages with positive input pulses.

The ordinary plate-coupled video amplifier is not suited to such cascaded service because it reverses the polarity of the applied signal. One alternative is to employ a cathode-coupled circuit that does not reverse the polarity. Two such circuits are the "bootstrap" circuit developed by Street, and the conventional cathode-coupled or "cathode-follower" circuit.

FIG. 235.—The "boot-strap" cathode-coupled amplifier.

The bootstrap circuit has its load resistance connected in the cathode circuit. The output is taken directly from the cathode terminal of the tube. The basic diagram is shown in Fig. 235. When a positive pulse is applied to the grid of such a stage, the tube and all associated elements connected to it are raised in potential by the increased voltage drop across the cathode load resistance. Coincidentally a positive pulse appears between the cathode terminal and ground. The name "bootstrap" was suggested by the self-raising action.

The design of the circuit is similar to that of a conventional video amplifier, except that the shunt capacitance existing across the output includes the capacitances of the tube structure and associated elements. The magnitude of this capacitance can be reduced by isolating and shielding the tube and associated elements and by the use of a filament-heating transformer with low capacitance between windings. When these precautions are taken, the bootstrap circuit can readily reproduce a pulse with a 0.1-μsec time of rise.

The essential feature of the bootstrap circuit, distinguishing it from the cathode-follower circuit, is the fact that the grid signal is referred to cathode potential. When conduction occurs in response to the positive applied pulse, the cathode rises from

its negative bias E_{kk} by an amount equal to the plate current times the load resistance. The plate potential does not change during the pulse. The stage has a gain exceeding unity. In fact, the gain has the same value as that of a conventional plate-coupled video amplifier employing the same tube and load resistance.

Since the bootstrap is used most widely as a high-level pulse amplifier (as a driver of the modulator tubes), it is often operated in nonlinear fashion. For this reason, its operation is best described in graphical terms. Figure 236 shows the operation of the circuit.

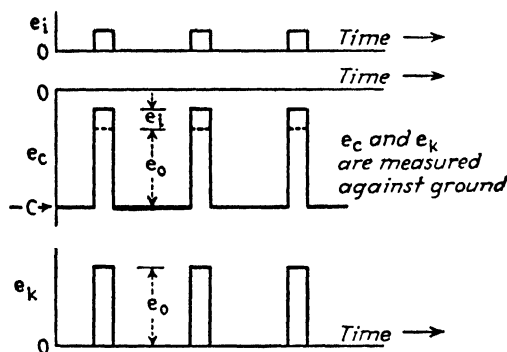



FIG. 236.—Voltage distribution of pulse-excited boot-strap circuit.

139. b. The Cathode-follower Circuit.—The conventional cathode-coupled stage or cathode follower (Fig. 237) differs from the bootstrap circuit in that the grid resistor is returned to ground rather than to the cathode. The voltage gain is always less than unity, because the output voltage across the cathode resistor is referred to the same level as the grid signal and is fed back to the grid circuit in degenerative polarity. Hence the cathode follower cannot serve as a voltage amplifier, but it is an efficient current amplifier. Stated in other words, the circuit has a low output impedance and is particularly suited to the feeding of coaxial cables and other low impedance circuits. In this respect it resembles a pulse transformer, but has the additional property of permitting direct coupling to the input circuit so that the l-f content (including the d-c reference) of the signal is preserved. The circuit has a high input impedance and excellent h-f response.

These properties can be deduced from consideration of the equivalent circuit (Fig. 237*a*). There are two equivalent generators in series, the applied voltage times the amplification factor μe_i and the degenerative voltage fed back from the cathode resistor $\mu i_p R_k$. The plate current resulting from the applied signal is, then

$$i_p = \frac{\mu e_i - \mu i_p R_k}{r_p + R_k} = \frac{\mu e_i}{r_p + R_k(\mu + 1)} \tag{446}$$

and the output voltage is $i_p R_k$. The gain ($G = i_p R_k / e_i$) is

$$G = \frac{\mu R_k}{r_p + R_k(\mu + 1)} \tag{447}$$

which is evidently smaller than unity. The gain can be held above 0.9, when the value of μ is higher than 1,000 and of r_p above 500,000 ohms (typical of pentode tubes).

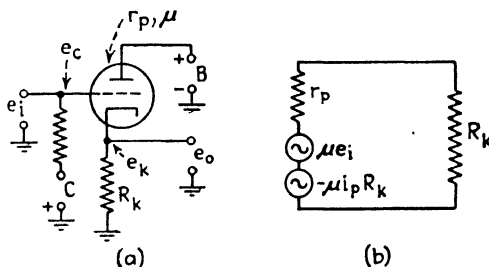



FIG. 237.—Cathode-follower amplifier, actual and equivalent circuits.

The cathode-coupled stage operates as if it were a conventional plate-coupled stage with an amplification factor of $\mu/(\mu + 1)$ and an internal plate resistance of $r_p/(\mu + 1)$ ohms. The internal plate resistance acts in parallel with the cathode resistance, producing an output impedance of

$$R_0 = \frac{R_k r_p/(\mu + 1)}{R_k + r_p/(\mu + 1)} = \frac{R_k r_p}{r_p + R_k(\mu + 1)} \tag{448}$$

In a typical case R_0 is 100 ohms (using the 6AC7 tube, $\mu = 6{,}750$, $r_p = 750{,}000$ ohms, with $R_k = 1{,}000$ ohms). Such a value is evidently suitable for feeding low-impedance circuits. The circuit is also an excellent means of feeding a highly capacitive circuit without loss of h-f response. In fact, the h-f limitations of the cathode follower reside wholly in the grid circuit. The fact that one end of the output circuit is at ground potential

permits direct connection with the grounded sheath of coaxial cables.

The circuit also possesses several of the advantages of the degenerative feedback amplifier, of which it is a particularly simple form. Among these advantages are stabilization of gain with respect to supply voltage changes and aging of tubes.

140. Pulse Transformers.—A device suitable for increasing the voltage or current level of a pulse sequence without materially changing in its shape is the pulse transformer. Pulse transformers thus serve the purposes of a video amplifier, except for power amplification. They are used to match pulse circuits to coaxial

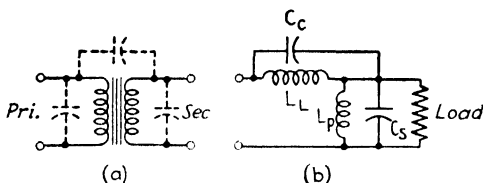



FIG. 238. Actual and equivalent circuits of a pulse transformer.

lines, to isolate direct current circuits, and to reverse pulse polarity. Prior to the development of radar equipment, iron-cored transformers capable of reproducing extremely short pulses were not generally available. Now they are available in great variety from small units working at peak levels of a few watts to large modulation transformers capable of handling 1-μsec pulses at a peak power level of several hundred thousand watts.

The action of the pulse transformer may be examined in terms of the equivalent circuit in Fig. 238. Two aspects are of interest: (1) the ability of the transformer to reproduce the leading and trailing edges of the pulse, and (2) its ability to maintain the amplitude of the flat top during the pulse. The transformer, as shown by the equivalent circuit, transmits the input signal to the load circuit through a series inductance L_L, the leakage inductance of the transformer. This inductance must be minimized if the high frequencies of the pulse spectrum are to be passed without loss. The leakage inductance may be reduced by adopting an interspersed construction in the primary and secondary of the transformer. One winding is wound within the other on the same leg of the core, so that the two occupy

as nearly as possible the same physical volume. If the number of turns in one winding is substantially less than that in the other, the physical size of the smaller winding may be increased by employing several coils in parallel. By such methods, the leakage inductance may be reduced to a few microhenries and leading edges of 0.1-μsec duration transformed without prolongation.

A more difficult problem is the maintenance of the flat top of the pulse. Here the essential element is the shunt inductance L_p, which is physically the inductance of the primary with the secondary open circuited. This inductance is related to the leakage inductance by

$$L_L = (1 - k)L_p \tag{449}$$

where k is the coefficient of coupling between the two windings. To make L_L small, for a given L_p, the coefficient of coupling must be as close to unity as possible. But L_p itself must not be too small, because the pulse current, acting through this inductance, produces the magnetization on which the maintenance of the flat top depends. To see the influence of L_p, suppose that the peak voltage of the applied pulse is V and its width d. Then the current i_m in L_p builds up during the pulse in accordance with

$$i_m = \frac{Vd}{L_p} \tag{450}$$

This current is subtracted from the signal current (passed through L_L). If i_m becomes appreciable during the pulse, the output current falls off during the flat top. Hence as large value of L_p as is consistent with the requirement for a small value of L_L is required.

A large value of L_p depends upon thorough magnetization of the iron core during the pulse. The iron must have a reasonably high value of permeability to insure adequate flux density. Otherwise the flux depends only upon proper mechanical design to insure full magnetization. Thin laminations must be used, and the resistivity of the core material must be high to inhibit eddy currents that would otherwise shield some of the material from magnetization. These requirements lead, incidentally, to high efficiency in the power transfer of the transformer, from 90 to 95 per cent.

When the trailing edge of the applied pulse appears, oscillations tend to appear in the secondary circuit, to dissipate the energy stored in L_p during the pulse. These oscillations are superimposed on the pulse and tend to prolong the trailing edge at the base. The base oscillations may be inhibited by a damping rectifier. Considerable care and ingenuity are required in the design of a pulse transformer for high-power applications, since the voltage breakdown limit between primary and secondary is inherently opposed to the requirement for close coupling between the windings, and the end-to-end insulation requirements in the high-voltage coil must also be satisfied.

141. Time Delay Circuits. a. Phase Shift of Sinusoidal Sources.—The use of sinusoidal timing sources in radar is pre-

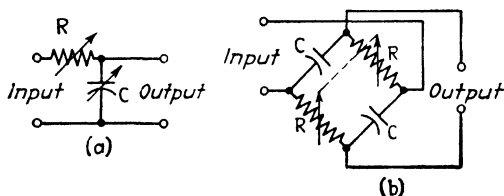



FIG. 239.—Phase shifters: single RC type (a), limited to 90°, and bridge type (b).

dicated in part on the fact that time delay may be introduced simply by shifting the phase of a sine wave. There are two types of shifters in use: those which shift the phase over a limited range, less than 360 deg; and the continuous phase-shift generator which generates additional a-c cycles and hence is capable of introducing any required amount of phase shift. Two simple forms of the first type are illustrated in Fig. 239. The single RC circuit is limited to a phase shift of less than 90 deg; the bridge-type RC shifter is limited to less than 180 deg.

When the phase shift is limited to less than a full cycle, a choice must be made between the amount of the time delay and the precision of the delay. Thus, if the period of the sine wave corresponds with the pulse interval, an adjustable time delay equal to approximately half the pulse interval may be obtained with the phase shifter shown in Fig. 239b. But, since the precision of setting the shifter is limited to the order of one electrical degree, the timing is accurate only to about 0.3 per cent of the pulse interval. If higher precision is required, a h-f wave may be used. Suppose the tenth harmonic of the basic

sine wave is used. This improves the precision ten times, but reduces the available time delay by the same factor.

The extent of the time delay may be increased without limit by employing a generator phase shifter, two forms of which are shown in Fig. 240. The Helmholtz coil (Fig. 240*a*) is essentially an induction generator, consisting of two stator coils which produce a rotating magnetic field and a rotor coil in which alternating voltage is induced by the field. The position of the rotor coil establishes the phase of the rotor voltage relative to the voltage across the stator. As the rotor is turned the phase is advanced or retarded (depending on the direction of

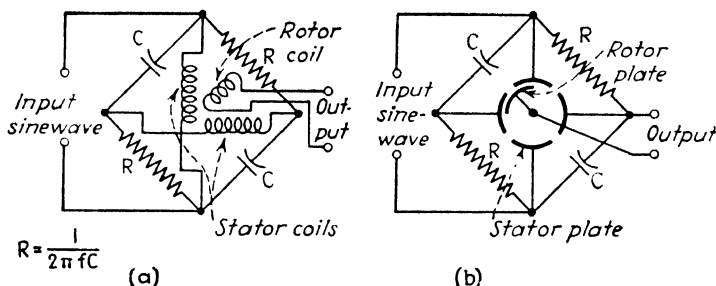



FIG. 240.—Generator-type phase shifters: (*a*) Helmholtz coil, (*b*) quadrantal capacitor.

rotation) 360 deg for each full revolution of the rotor. The accuracy of reading the phase setting (1 deg accuracy is not difficult to achieve) remains substantially independent of the frequency used, so it is possible to measure the time delay very precisely by the use of high frequencies. Moreover, any amount of time delay may be introduced by continued rotation of the rotor, and thus the high precision may be extended over long periods of time. It is, of course, necessary to start the phase shift from zero when the measurement is commenced and to count the cycles thereafter (or otherwise to identify the particular cycle on which the timing is based).

The capacitor phase shifter, shown in Fig. 240*b*, serves the same purpose and is somewhat easier to construct in precise form. Each of the four stator quadrants of the capacitor is fed voltage 90 deg out of phase with the adjacent quadrant by the associated resistor-capacitor network. The rotor plate assumes a voltage whose phase is determined by the vector sum of the voltages

induced by the four quadrants, and the phase is adjustable by rotation throughout any portion or multiple of 360 deg. The plates are shaped to maintain the amplitude of the output voltage constant at all positions of the rotor and to assure strict proportionality between the mechanical position of the rotor and the corresponding electrical phase shift. The output voltage (which may be taken between the rotor and any one of the stator quadrants) is generally fed to a high impedance amplifier to avoid the effects of resistance loading. The method of application is otherwise identical with that of the Helmholtz coil.

142. b. Artificial Transmission Lines.—A pulse sequence may be delayed by passing it through a transmission line. Real transmission lines (for example, of the coaxial variety) introduce

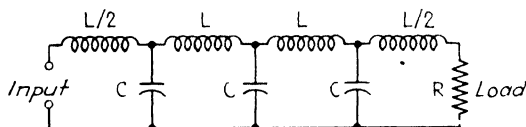



FIG. 241 Constant-K artificial transmission line.

a delay of about $^1\!/_{10}$ μsec per 100 ft. A larger delay can be achieved in a smaller space by the use of an artificial transmission line or low-pass filter. Delays up to 100 μsec may be accomplished in such circuits, provided that the associated attenuation and distortion of the pulse shape can be tolerated.

The design of delay lines is illustrated in the constant-K low-pass filter shown in Fig. 241. The time delay introduced by each section of the line is

$$t = \sqrt{LC} \quad \mu\text{secs} \tag{451}$$

where L is in microhenries and C is in microfarads. A long delay can be obtained by employing many sections or by employing large values of L and C. Unfortunately, neither of these methods can be pursued without limit if short pulses are to be delayed without excessive attenuation. A sharp pulse implies a wide spectrum and correspondingly a high cutoff frequency in the filter network. The nominal cutoff frequency of a constant-K filter is given by

$$f_c = \frac{R}{\pi L} = \frac{1}{\pi C R} \tag{452}$$

Hence small values of L and C are mandatory if a sharp pulse is to be passed without attenuation of the higher frequencies in its spectrum (and consequent distortion of its shape). This is in direct opposition to the requirement for a long delay in each section. The alternative, the use of many sections, reaches a limit when the attenuation within the passband becomes excessive. Other filter structures have better properties than the constant K. One of the best employs negative mutual inductance between coils in adjacent sections. But even this circuit introduces difficulties when it is attempted to delay a 1-μsec pulse for a period longer than 100 μsec.

Another form of delay line, operating on similar principles, employs the transmission of supersonic waves through a liquid.

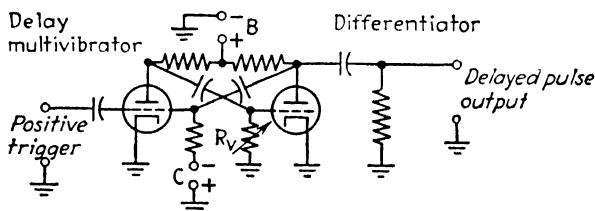



FIG. 242.—Delay multivibrator and differentiator.

The delay of wave propagation through mercury, for example, is about 18 μsec per in., at room temperature, or about 700 μsec per m. The mercury is column excited at one end by a piezoelectric crystal, which is operated at a carrier frequency, the carrier being modulated with the pulse waveform. The opposite end of the column is fitted with a similar crystal that transforms the supersonic wave into an electrical signal. In practice great care must be taken in the design of a supersonic delay line to avoid excessive attenuation and unwanted reflections.

143. c. Delay Multivibrators.—A simple method of introducing an adjustable time delay, with reasonable accuracy and stability, is the delay multivibrator illustrated in Fig. 242. This circuit is in no way different from the unsymmetrical biased multivibrator previously discussed (Secs. 123 and 127).

The tube with the shorter conduction period (at left in Fig. 242) is biased to cutoff and the signal pulse to be delayed is applied to it in positive polarity. The circuit thereupon reacts, driving the opposite tube to cutoff, in which condition it remains while its grid (with a large and controllable time constant)

recovers from the cutoff condition. When it does so, the grid of the first tube falls below cutoff forming the trailing edge of a rectangular wave. This trailing edge follows the initial synchronizing pulse by a time dependent on the time constant R_2C_2, which may be of any required length from a few microseconds to several seconds, and which may be adjusted by variation of R_2.

The trailing edge of the long rectangular wave is passed through a differentiator amplifier (Fig. 242) in the output of which appears a positive pulse bearing a fixed time relationship to the positive pulse applied to the multivibrator. The precision of the timing in this circuit is proportional to the total delay introduced, and is dependent on the constancy with which the values of the circuit elements, tube characteristics and applied voltages are maintained.

144. Pulse Circuit Auxiliaries. a. Maintenance of Direct-current Reference.—In a number of applications (notably the

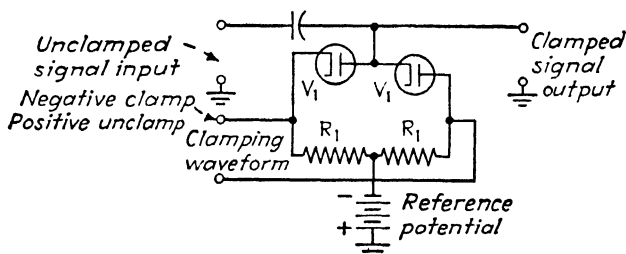



FIG. 243.—Diode clamping circuit.

maintenance of background indicator brightness, and fixing the center of deflection in plan position indicators), it is necessary to assign a particular value of current or voltage to a particular level in a waveform. Whenever a waveform is passed through capacitive coupling or through a transformer, the average ordinate of the waveform assumes zero potential, since neither device passes direct current. If it is desired that some other ordinate (for example, the base of the sawtooth current wave used to cause radial deflection in a plan position indicator) assume zero level, auxiliary means must be taken to introduce the change. Such auxiliary circuits are known as "d-c inserters" or "clamping circuits." An elementary example of the clamping circuit is shown in Fig. 243. A rectangular control waveform

is applied to the diodes in synchronism with the signal waveform to be shifted. During the active portion of the signal wave, the positive half of the clamping wave is applied, the diodes do not conduct, and the output of the circuit is free to follow the signal applied to it. During the inactive (intervening) time, the negative half of the synchronizing wave is applied, causing the diodes to conduct and connect the signal circuit with the desired potential. The output waveform is thus "clamped" to the reference potential until the end of the negative synchronizing wave, whereupon the circuit is again free to follow the applied signal waveform. In other applications (for example, the automatic control of background brightness in an indicator) it is desired to fix some level in a waveform irrespective of changes in the average value of the waveform. Usually it is permissible to fix the peak value of the waveform. This may be done (as shown in Fig. 244) by passing the waveform through a diode to a capacitor. The combination acts as a peak detector, the direct voltage across the capacitor assuming (approximately) the peak value of the waveform. A bias voltage interposed in series with the capacitor voltage serves to set the d-c level at any required value, at which value it remains regardless of the average value of the waveform itself. This circuit is useful in preventing noise, or stray pulses from other radars, from affecting the brightness of the cathode-ray indicator.

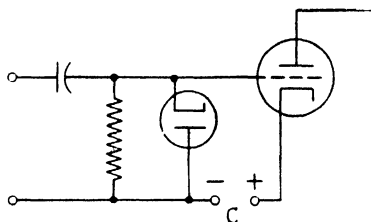



FIG. 244.—Diode d-c restorer.

145. b. Regulated Power Supply.—The common power supply used in pulse systems requires particular attention because of the heavy and irregular demands put upon it. Whenever several pulse sequences are to be derived simultaneously, care must be taken to avoid reaction of one sequence upon the other through voltage variations in the power supply. For this and other reasons, regulated power supplies are commonly employed in pulse systems. Regulators of the gaseous discharge type serve when the current demands are not too severe. A circuit generally used for heavy duty is that shown in Fig. 245. The output current passes through the high-current triode tubes (usually

several in parallel). The grid voltage of the series triodes is controlled by the d-c amplifier tube, which operates on variations in the output voltage. The gas-filled tube serves as a stable cathode bias source for the d-c amplifier.

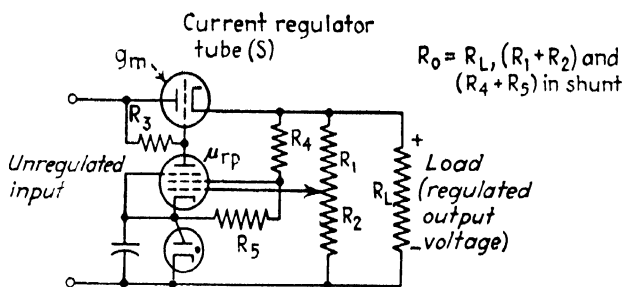



FIG. 245.—Series triode voltage regulator.

The circuit is adjusted by the tap on R_2 until the following relation holds

$$\frac{\mu g_m R_2 R_3 R_0}{(R_1 + R_2)(R_3 + r_p)} = 1 \tag{453}$$

where μ, r_p, and g_m apply to the series triode tubes as a group, the resistances are as shown, and R_0 is the total shunt resistance composed of R_L shunted by $R_1 + R_2$ and $R_4 + R_5$. Larger values of R_2 cause the circuit to overcompensate and hunt about the steady state, whereas smaller values cause incomplete suppression of voltage variations in the output. The balance condition is most easily found by noting the change in output voltage as the load resistance is changed and by adjusting R_2 until the voltage variation is a minimum.

CHAPTER VII

BASIC RADIO-FREQUENCY CIRCUITS AND STRUCTURES

The r-f (radio-frequency) functions of a radar system are centered in several circuits and structures that are common to all equipments. Listed in order of the initial passage of the signal through them, these basic r-f devices are the r-f transmitting oscillator, the duplex switching system, the radiator, the receiver r-f amplifier, and the mixer (including local oscillator). These elements are connected by a transmission system (coaxial line or waveguide) that includes a number of specialized fittings, joints, tuning members, and terminations. In this chapter we discuss the operations of these r-f devices.

146. Coaxial Lines, Waveguides, and Associated Fittings.—The interconnecting lines or guides of a radar system are required to fulfill three functions: (1) to pass power at the required power level without excessive loss, (2) to match impedances of the connected devices, and (3) to introduce mechanical degrees of freedom, particularly those of rotation or reciprocation about the scanning axes. These functions must be fulfilled over as wide a band of frequency as possible, to permit the shifting of the transmitter frequency without retuning the transmission system.

In Chap. III the theory of coaxial lines and waveguides and their design parameters has been discussed at length. Certain practical aspects of transmission systems, particularly the mechanical construction of the lines and associated fittings, are treated in the following sections.

147. Support of Coaxial Inner Conductors.—The insulation of the center conductor of a coaxial cable must be of high quality to avoid excessive losses. At frequencies below 3,000 megacycles solid dielectrics (polyethylene, or polyisobutylene compounded with cyclized rubber or polystyrene) are commonly employed. In the microwave region these materials display appreciable losses (0.2 to 0.3 db per ft at 3,000 megacycles) so they are not used except in short lengths to introduce necessary flexibility.

Flexibility may also be achieved in beaded cable, in which the inner conductor is supported by polystyrene beads. The signal reflections that occur at each bead may be reduced by spacing the beads in groups so that the reflections tend to cancel each other. The system may be compounded by proper spacing of the bead groups to cancel reflections between groups. This is the "Lawson line" that was used in the early microwave equipments. Beaded cables have been supplanted by nonflexible stub-supported lines combined with short lengths of solid dielectric coaxial cable.

The most widely used r-f coaxial line is the stub-supported line illustrated in Fig. 246*a*. The inner conductor is supported

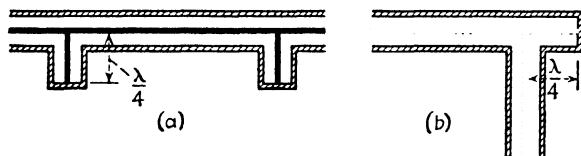



FIG. 246.—Coaxial stub supports.

at intervals by quarter-wave sections of line, each short-circuited at the outer end. The impedance presented by the stub at the inner conductor is theoretically infinite, and therefore the inner conductor is effectively "insulated." The quarter-wave stub may also be used to introduce a right angle bend, as illustrated in Fig. 246*b*.

The simple quarter-wave stub presents high impedance only in the near vicinity of the frequency at which its length is a quarter wavelength. Since changes in frequency must often be accommodated in a radar system, means must be found to "broadband" the stub-supported line so that it will operate over a wide frequency range.

The technique usually adopted is illustrated in Fig. 247. The diameter of the center conductor is increased for a distance of approximately a half wavelength. The stub support, placed at the center of this region, is also enlarged. The enlarged center conductor acts as a half-wave transformer. At the design frequency (that at which the stub is a quarter-wave long) the half-wave transformer has no effect since it introduces a one-to-one impedance transformation that matches the two segments of line. Off the design frequency, the stub becomes a reactive

element, and the half-wave segment is no longer a one-to-one transformer. However the two reactances combine to present a resistive impedance at the terminations (*a* and *b* in the figure).

Figure 247 shows the circle diagrams of the broadband stub above and below the design frequency. The reactive effect

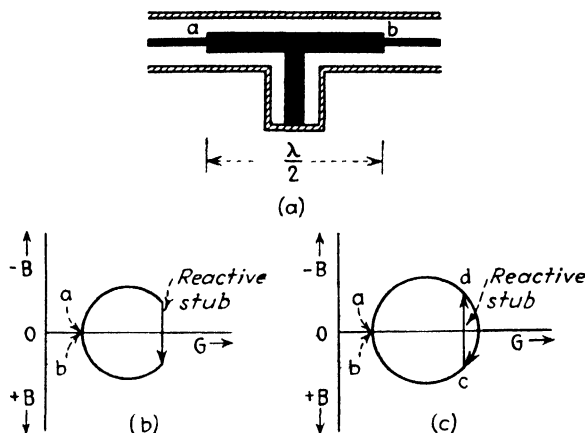



FIG. 247.- Broad-band stub support (*a*), and circle diagrams below (*b*) and above (*c*) the resonant frequency

of the stub, in the off-frequency cases, completes the circle. Thus the impedances at points *a* and *b* are resistive and equal to the characteristic impedance, and no reflections occur. In practice such broadband stub-supported lines can cover a 15 per cent variation in frequency (450 megacycles at 3,000 megacycles) with standing-wave ratio (in voltage) less than 1.01.

A universal stub-supported right angle bend is shown in Fig. 248. Generally the characteristic admittance of the side member is somewhat larger than that of the straight-through member. Universal shorting plugs and connectors are provided so that any one of the openings shown may be plugged and the remaining two connected to the lines.

FIG. 248.—Universal three-way joint.

When two stub-supported lines are connected, care must be taken to minimize reflection losses at the joint. Figure 249 shows a typical "bullet and flange" joint, designed to present a

smooth, continuous surface across the joint along inner and outer conductors. The inner conductor is joined by a bullet-shaped member that makes a pressure fit in each inner conductor. The flanges of the outer conductor are tapered so that when drawn together they make firm contact. A ⅞-in. 50-ohm line joined in this manner will carry pulses at a peak power of 200 kw reliably. A 1⅝-in. line of the same impedance (the largest feasible size for 10-cm energy if higher modes are to be avoided) will carry 1,000 kw.

148. Rotating Joints in Coaxial Lines.—When continuous rotary scanning (circular, helical, conical, etc.) is employed,

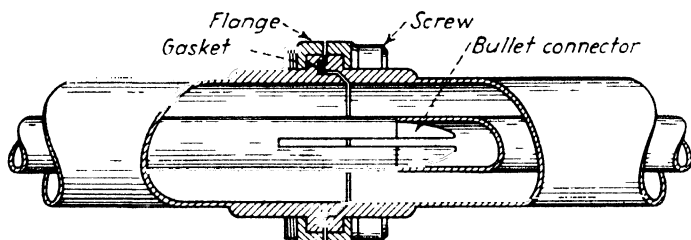



FIG. 249—Bullet and flange straight-through joint.

one or more continuously rotating joints are required in the transmission line between transmitter and radiator. When a reciprocating scan is used, such rotating joints may be used or a flexible coil of cable may be employed to permit motion over the required sector.

The axial symmetry of the principal mode in coaxial cable lends itself to the design of a rotating joint, but practical problems arise if the degree of contact between the two members of the joint is not constant. Actual physical contact between the members is possible, but impractical because of the difficulty of maintaining the contact resistance constant over an extended period. The solution usually adopted is to employ capacitive coupling between the members of the joint and to prevent leakage of energy through the capacitive opening by quarter-wave choke segments.

A typical rotating joint of this type is illustrated in Fig. 250. The clearance between adjacent surfaces is made as small as possible, within the mechanical tolerances of the assembly, in the interest of high capacitance between the members The length of the overlap is generally taken as a quarter wavelength

or somewhat less. Longer overlaps introduce inductive reactance and impede the transfer of energy. When the overlap is exactly a quarter wavelength, the radiation resistance of the outer edge of the capacitive gap is transformed to zero impedance at the inner edge, regardless of the clearance between the surfaces. A

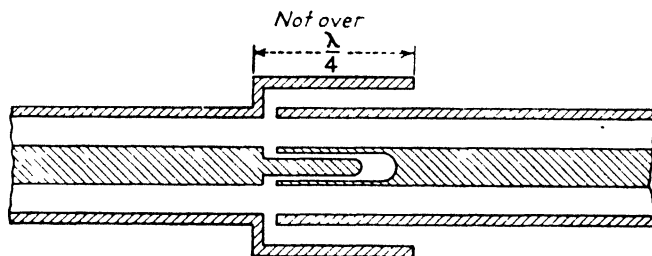



FIG. 250.—Simple capacitive coaxial rotating joint

small clearance helps in maintaining low impedance across the gap at frequencies off the design value.

To suppress further the radiation through the gap, an inductive "choke" element may be placed in series with the radiation resistance, as shown in Fig. 251. The choke is a quarter-wave

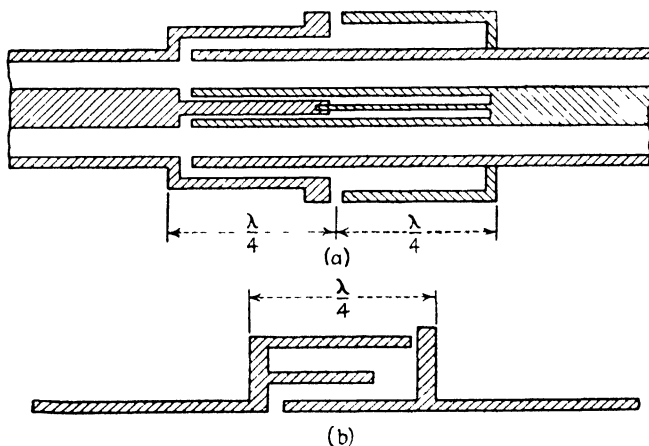



FIG. 251.—Choke flange coaxial joints, extended (*a*) and folded (*b*).

segment short-circuited at one end, the outer surface of the coaxial line serving as the inner conductor. Because the high impedance of the choke is in series with the radiation resistance, imperfect electrical contact at the faces (Fig. 251*a*) does not introduce adverse effects. A different physical form appears

in Fig. 251*b*. Here the series choke element is in effect folded over and placed above the capacitive element, thus reducing the length of the joint. Other variations of the rotating flange assembly (some including dielectric substances to shorten the joint) are in use. The standing-wave ratio of a well-designed rotating joint is of the order of 1.04 to 1.08. Since the concentration of electric field at the gap may produce sparkover, the rotating joint is often the limiting factor in the power-handling capacity of the transmission line system.

149. Waveguide Joints.—The design of waveguides and their tuning elements (apertures and screws) is treated in Chap. III, to which the reader is referred. The principal practical problems

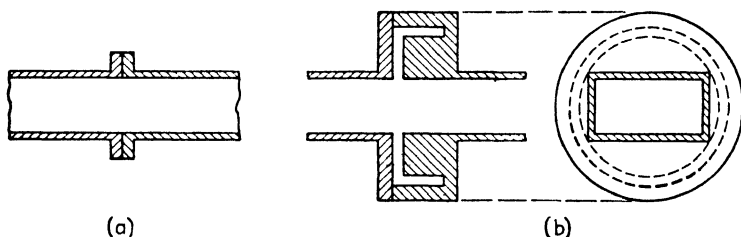



FIG. 252. Butt waveguide joint (*a*) and choke joint (*b*).

are the introduction of flexible or rotating joints. The simplest joint is the nonflexible form shown in Fig. 252*a*. This joint consists of two butt flanges held together by screws. The screws are accurately positioned so that the inner surface of the guide is smooth across the joint.

When some degree of flexibility is required, the guide is commonly joined in a circular flange, into which is cut a quarter-wave choke, as shown in Fig. 252*b*. The circular form is employed in the flange, even when coupling rectangular guides, to avoid end effects at the corners of the joint cross section. If a considerable degree of flexibility is required, the two members of the joint may be separated by a small fraction of wavelength, without excessive radiation. A separation of 0.06 wavelength leads to a standing-wave ratio of less than 1.3 and a power loss of about 0.3 db. This compares with about 0.02 db for a rigid choke-flange connection and 0.05 db for a simple contact joint.

Right-angle joints in waveguide are generally in the form of "tee" sections (see Chap. III, page 203) with a tuning plunger and diaphragms to cancel reflections. When matching dia-

phragms are used, the standing-wave ratio is generally below 1.05, whereas without them the ratio is generally 1.3 or higher.

Rotating joints in waveguide must, of course, be based on an axially symmetrical mode, of which the $TM_{0,1}$ mode is the lowest form. Since this mode is not widely used in ordinary practice (the $TE_{1,0}$ mode in rectangular guide is preferred because of lower losses due to mismatching), it is generally necessary to transform from the $TE_{1,0}$ rectangular mode to the $TM_{0,1}$ circular

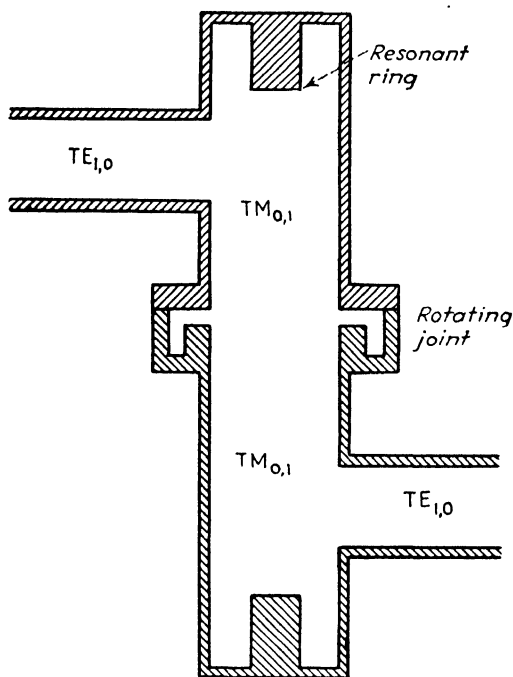



FIG. 253.—Waveguide rotating joint, circular elements connecting rectangular guides.

mode and back again. A typical transforming section of this type is shown in Fig. 253. The rotating joint itself is similar to that used in the outer conductor of a coaxial rotating joint (Fig. 251). A limitation of this type of rotating joint is the distance between the rectangular guides (approximately the length of the circular guide). This distance should not be greater than about 10 wavelengths, to insure proper operation over a band equal to 2 per cent of the carrier value.

When waveguides are used as the basic transmission medium, it is often desirable to include coaxial elements in certain parts of the system. For example, the cavity magnetron output is usually of the coaxial varity, because energy is most efficiently abstracted from the cavity by a coupling loop. Also the coaxial form is particularly convenient in rotating joints. Hence elements used to join coaxial and waveguide segments assume practical importance. Two widely used methods of connecting coaxial lines and waveguides are the probe and stub feeds illustrated in Fig. 254.

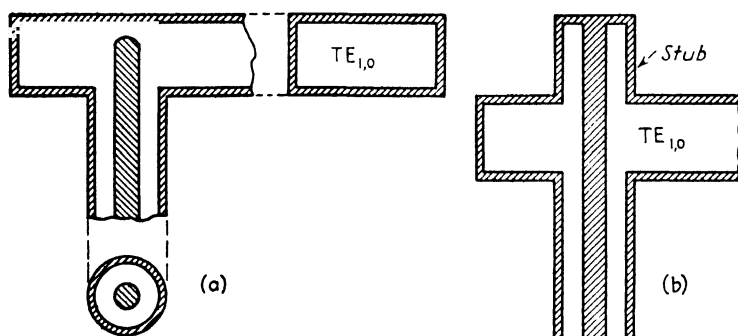



FIG. 254.—Coaxial-waveguide connectors: (*a*) probe type, (*b*) stub-supported type.

150. Radio-frequency Generators (Transmitting Oscillators). Radar r-f generators, or transmitting oscillators, are basically the same as the oscillators used at lower frequencies. Each employs a modulated electron stream, which passes between two oscillating systems (circuits, lines, or cavities). The oscillating systems are coupled in such a way that oscillations in one system are sustained by periodic contributions of energy from the other. In uhf (ultrahigh-frequency) and shf (superhigh-frequency) practice, the oscillating systems must be closely integrated with the electronic element to achieve high-power output and high efficiency and the electronic element must be designed to sustain a high energy level during each pulse.

The physical form of the transmitting oscillator depends on the frequency of operation. Triodes, constructed for uhf operation, are used from 100 megacycles (the lower limit of the radar spectrum) to 3,000 megacycles. High peak-power output (100 kw and above) cannot be achieved with triodes at fre-

quencies much above 600 megacycles.[1] For high-power service at frequencies from 600 megacycles to the present upper limit of the radar spectrum (30,000 megacycles), the tube universally used is the cavity magnetron. High-power pulsed operation at frequencies above 500 megacycles has also been achieved in klystron tubes, particularly by British workers.

The form of the tuned circuits varies also with the frequency. At frequencies from 100 to 600 megacycles, external transmission-line segments, usually of the parallel conductor type, are used. In one oscillator, used at 500 to 600 megacycles, the line segments are built into the tube structure, within the vacuum envelope. Triodes used above 500 megacycles are integrated with resonant cavities, which are usually of the coaxial type. Similarly, magnetrons used above 500 megacycles employ resonant cavities, of cylindrical or slot-like form, in multiple arrays of six or more.

High-power output is achieved by providing a cathode capable of high peak emission and by drawing the electrons from the cathode at a high anode voltage. In triodes the high anode voltage minimizes transit time difficulties, and in all types of tube it permits correspondingly high-power output from a given electron emitter.

The structural design of uhf oscillators must minimize electrode capacitance and series inductance in the leads. At the high frequencies, 3,000 megacycles and above, the output energy is abstracted directly from the cavity by coaxial coupling loops or waveguide apertures, which, when properly matched, introduce no capacitive or inductive effects.

151. External-line Triode Oscillators.—Several uhf oscillator triodes designed for external resonant lines are shown in Figs. 255, 256, and 257. The tubes most nearly resembling conventional l-f (low-frequency) tubes are those contained in glass envelopes (Fig. 255). Two leads are connected to grid and to plate to reduce the lead inductance. The filament and plate leads are brought out at opposite ends of the tube to reduce capacitance, as well as to introduce maximum insulation against the high anode voltage (up to 25 kv). The grid is a cage of axial wires, rather than a spiral, in order to reduce the electrode

[1] LAW, R. R., and others, Development of pulse triodes and circuits to give one megawatt at 600 mc, *R.C.A. Rev.*, **7**, 253 (June, 1946).

inductance. The filaments are generally of thoriated tungsten or tantalum, since these are materials capable of standing up

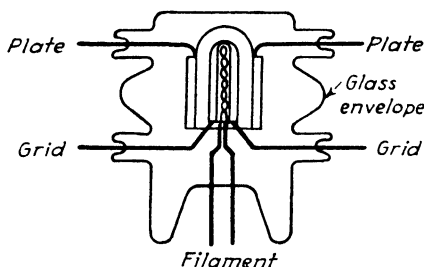



FIG. 255.—Early form of *v-h-f* pulse-type triode. Double leads on plate and grid reduce lead inductance.

under high anode voltages. Since the emission efficiency of tungsten is low (about 50 ma per watt) the tubes require very large heating currents. Thus to supply a peak emission of 10 amp per tube (typical in a two-tube 200-kw oscillator at 15 kv), nearly 400 watts of filament heating power are required. This heating power is supplied at low voltage, since the filament must have a large surface area and consequently has low resistance.

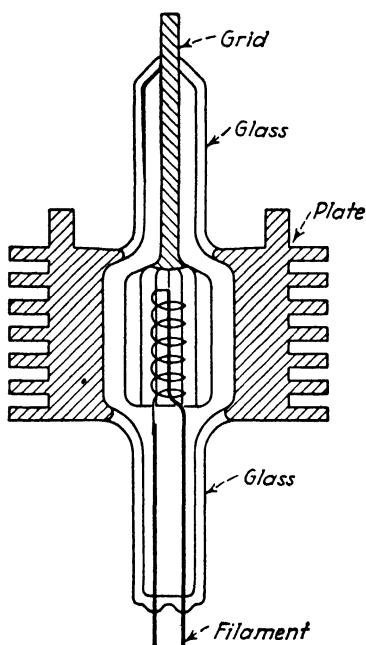



FIG. 256.—Micropup pulse-type triode.

Another triode construction, which originated in England, is that of the so-called "micropup" tube (Fig. 256). In this tube the body of the tube is the anode, with fins attached for air cooling. The tungsten filament enters through a glass segment at one end and the grid through the other.

A final form of tube (Fig. 257) is the disk-seal tube,[1] intended for use with resonant cavities. In

[1] McARTHUR and PETERSON: The Lighthouse Tube, *Proc. National Electronics Conference*, p. 38 (Chicago, 1944). JAMIESON and WHINNERY: Power Amplifiers with Disk-seal Tubes, *Proc. I.R.E.*, **34,** 483 (July, 1946).

this tube, the anode, grid, and cathode are brought out to disks separated by glass cylinders. The tube is inserted in a cavity structure and the disks form an integral part of the resonators. The center posts of the tube (the anode and cathode) form part of the cavity. Tuning is accomplished by deforming the cavity, by inserting tuning plugs, or by adjusting the length of the cavity.

FIG. 257.—Disk-seal triode ("lighthouse tube").

The tuned circuits used with external-circuit tubes are segments of parallel transmission lines, tuned at or near a quarter wavelength. A typical form is shown in Fig. 258. The line conductors are joined by a yoke at the far end, to the center of which the d-c potential is applied. Another member, called the "shorting bar," slides along the line, tuning the line to the proper value. The conductors are often made of steel tubing for mechanical strength and rigidity. Since the current flow is confined by skin effect to a thin section at the surface of the conductor, losses can be minimized by reducing the surface resistance. A plating of a highly conducting metal (usually silver) on large-diameter conductors is generally employed.

FIG. 258.—Parallel-line tuned circuit.

As pointed out in Chap. III, a short-circuited transmission line displays infinite input impedance when a quarter wavelength long. This effect can be used, as in the stub-supported line, to provide r-f isolation. When the line is shorter than a quarter wavelength it displays an inductive reactance; when longer,

capacitive reactance. These reactances are introduced when required to satisfy the condition of oscillation and to establish the desired frequency of oscillation.

The electrical length of resonant lines can be changed by other means than by sliding the shorting bar. The electrical length is increased when the conductors are spread apart at the shorted end, and it is decreased when they are brought together. Near-by shielding also may affect the electrical length, although the effect is small if the shield is disposed symmetrically with respect to the conductors. Lumped reactances, in the form of auxiliary capacitor plates, are often combined with the resonant lines. Such lumped capacitance increases the electrical length of the line.

An important aspect of uhf oscillator design is the suppression of unwanted modes of oscillation. In push-pull circuits and in multitube ring oscillators, there are generally several alternate electrical paths between the elements of the tubes. Thus a push-pull oscillator may tend to oscillate with the tubes effectively in parallel. Such oscillations not only waste power but interfere with the normal oscillation and affect the shape of the desired pulse. To eliminate such parasitic oscillations, two methods may be used: One is to adjust the circuit so that the phase relation between grid and plate over the parasitic path cannot support oscillations. The other, more commonly used, is to insert resistance to absorb energy, or chokes to prevent its passage, along the parasitic path. Small noninductive resistors directly at the grids of a push-pull oscillator are often used as suppressors. Resistors or chokes in series with the plate supply and the grid supply in a push-pull oscillator serve to inhibit parallel oscillations while not affecting the normal push-pull mode of oscillation.

152. Ultrahigh-frequency Oscillator Circuits.—The circuits used in the uhf region (100 to 600 megacycles) are designated by the number of tubes employed and by the location of the tuned circuits. For low-power service a single tube may suffice. The disk-seal types are generally used in single-tube circuits, because the resonant cavity construction does not lend itself to multitube arrangements.

By far the majority of triode radar transmitters use two or more tubes in push-pull or ring circuits. In these circuits, the tube and circuit capacitances are arranged in series, so that excessive

shunt capacitance can be avoided and a correspondingly high oscillation frequency achieved. The output coupling is arranged to derive power from each tube so that the power output is in direct proportion to the number of tubes used. Ring circuits containing as many as 16 tubes were used in early radar designs.

The designation of the circuit by the location of the tuned elements follows the terminology familiar in l-f applications. At least two tuned circuits are required, and they may be placed in the grid, plate, or cathode circuits. Two common arrangements are the tuned-plate tuned-grid and the tuned-grid tuned-cathode. Use is occasionally made of three circuits, as in the tuned-plate tuned-grid tuned-cathode oscillator.

FIG. 259.—Basic triode oscillators: (*a*) Hartley and (*b*) Colpitts.

These oscillators can be reduced to one of two equivalent circuits, shown in Fig. 259, regardless of the number of tubes and position of the oscillating circuits. These are the Hartley and Colpitts oscillators. The operation of oscillators depends on two conditions that are readily identified in the equivalent form. Consider the Colpitts equivalent circuit (Fig. 259*b*), and suppose that an r-f signal is applied in series with its grid, at the terminals shown. The signal is reproduced, in reverse polarity, in the plate circuit where it appears across the plate-cathode capacitance of the tube C_{pk}. The r-f current fed to C_{pk} excites a circulating current of much larger magnitude in the tank circuit. The circulating current, in turn, produces a voltage drop across the grid-cathode capacitance of the tube C_{gk}. The phase of this latter voltage is such as to drive the grid more positive as the plate becomes more negative, that is, it has the same phase as the initially assumed grid signal. Thus the induced grid voltage may be substituted for the signal. In other words, the oscillator is essentially an amplifier that supplies its own excitation.

The self-oscillation will continue if, and only if, the r-f amplification of the tube is sufficiently high to cause the fed-back grid voltage to equal or exceed the grid voltage originally present.

If the tube gain is G, and the two tube capacitances are C_{pk} and C_{gk}, the condition of oscillation is realized if

$$G \geqq \frac{C_{gk}}{C_{pk}} \tag{454}$$

Since, for a given tube and loading, the gain G is fixed, oscillations are induced only if the values C_{gk} and C_{pk} satisfy Eq. (454). If they do not, additional capacitance or inductance must be added at the appropriate point. In other words, the position of the cathode tap in the capacitance branch of the Colpitts circuit determines the amount of feedback.

Feedback voltage higher than the critical amount represented in Eq. (454) is desirable in the interest of stable oscillation and high efficiency. High efficiency is obtained by operating the oscillator as a class C amplifier, with sufficient grid bias to restrict the flow of plate current to periods substantially shorter than a half cycle. The grid bias may be supplied from a fixed source, but it is more usual to supply at least part of the bias from grid current, which charges the grid-cathode capacitance. A resistance ("grid leak") is connected across this capacitance (from grid to r-f ground). The desired value of grid-leak bias is obtained by adjusting the resistance until the discharge through it, at the desired voltage, just balances the charge from grid current.

When the circuit adjustment for oscillation at maximum efficiency has been found, the circuit must be further adjusted to produce the required frequency of oscillation. The oscillation frequency is determined by the shunt inductance and the capacitance of the tank circuit. In the Colpitts oscillator (Fig. 259*b*), the oscillation frequency is

$$f_0 = \frac{1}{2\pi \sqrt{L\left\{\left(\dfrac{C_{gk}C_{pk}}{C_{gk} + C_{pk}}\right) + C_{gp}\right\}}} \tag{455}$$

In theory the conditions for sufficient feedback, Eq. (454), and for the required frequency, Eq. (455), can be satisfied independently. This is true, however, only if the capacitances C_{gk} and C_{pk} are simple capacitances. In uhf oscillators, one or both of these is, in fact, a more complicated reactance, consisting of the tube capacitance plus a resonant line. Adjustment of this

combined reactance affects the inductive component of the tank circuit, as well as the ratio C_{gk}/C_{gp} of the capacitive voltage divider, and thus in general the adjustments for feedback and frequency are not independent. One advantage of the tuned-grid tuned-cathode circuit is that the two adjustments are more nearly independent than in other circuits.

The actual and equivalent circuits of the tuned-grid tuned-plate oscillator are given in Fig. 260. Both resonant lines are tuned inductively, that is, less than a quarter wave long. The tube capacitances C_{gk} and C_{pk} combine with the inductive lines

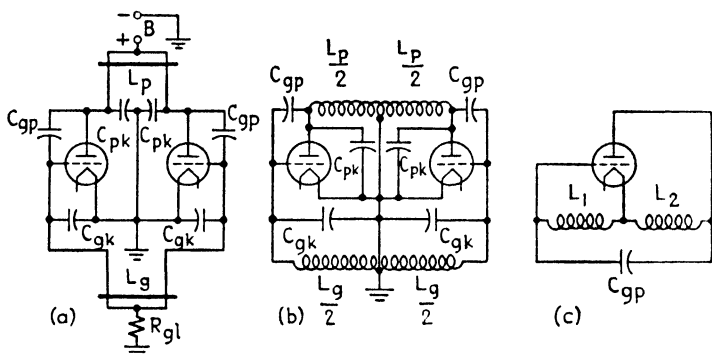



FIG. 260.—Tuned-plate tuned-grid line-circuit oscillator: (a) actual circuit, (b) two-tube equivalent, and (c) one-tube equivalent.

to form two tuned circuits, the outer ends of which are connected through the grid-plate capacitance C_{gp}. The inductance of the filament leads is cancelled by series capacitance, and therefore the cathode may be represented as connected to r-f ground.

One half of the push-pull circuit (that to the left of the ground plane) is taken as the equivalent circuit (Fig. 260c). The L_gC_{gk} and L_pC_{pk} branches are tuned to the inductive condition, so that the final equivalent circuit (Fig. 260c), is the Hartley circuit. Since the resonant lines are tuned to the inductive condition, their natural resonances are higher than the oscillating frequency. One line, that whose natural resonance is closest to the oscillating frequency, controls the frequency of oscillation. Tuning the other circuit affects the degree of feedback. Detuning the cathode series-resonant circuit also affects the feedback. There are, in other words, three interdependent adjustments in the tuned-plate tuned-grid circuit.

In the tuned-grid tuned-cathode circuit (Fig. 261) the plates of the two tubes are connected together by a noninductive short circuit. Each grid resonant line is tuned to the inductive condition, combining with the associated grid-plate tube capacitances as shown in Fig. 261*b*. One half of this circuit, to the left of the ground plane, forms the equivalent one-tube circuit. The cathode line is tuned to the oscillating frequency, thus presenting a resistance across the plate-cathode capacitance. The final equivalent circuit (Fig. 261*c*), is of the Colpitts type. Adjustment of the grid lines controls the oscillation frequency, while a

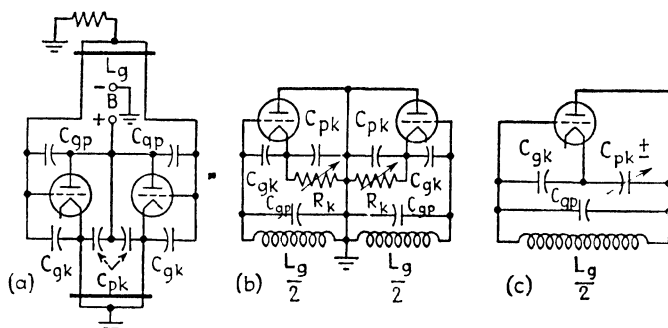



FIG. 261.—Tuned-grid tuned-cathode line-circuit oscillator: actual (*a*), two-tube (*b*), and one-tube (*c*) equivalent circuits.

change in the cathode line adds to or subtracts from the capacitance C_{pk}, thus varying the ratio of C_{gk}/C_{pk}, which controls the feedback. Since the cathode line is approximately resistive, it has little effect on the oscillation frequency, and the feedback and frequency adjustments are thus substantially separate. This fact, plus the fact that there are only two tuning controls (no tuning is required in the plate circuit) make this oscillator easier to adjust than the tuned-grid tuned-plate form.

The tuned-grid tuned-cathode circuit has the advantage, also, that the voltage on the anodes is simply the d-c value. Conversely no direct voltage is present in the cathode circuit, that is, the peak value is that of the radio frequency alone. This separation of the d-c and r-f peak potentials reduces the danger of insulation breakdown, corona, etc.

An important advantage of the parallel-line push-pull circuits just described is the fact that the center point of the yoke of each resonant line is inherently at r-f ground potential. This follows

from the inherent balance, that is, one conductor rises in potential by precisely the amount the other conductor falls. On this account no special precautions need be taken to establish r-f grounds, whereas in single-tube circuits chokes or auxiliary tuned lines are often necessary for this purpose.

An interesting special case of the push-pull tuned-plate tuned-grid oscillator, shown in Fig. 262, is the internal-circuit tube developed by Zahl. This tube contains four triodes connected directly to resonant quarter wave lines within the vacuum envelope. This tube can generate power of about 200 kw peak,

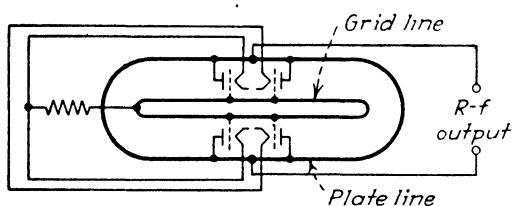



FIG. 262.—Internal-circuit tube oscillator.

and can be tuned over a limited range by adjustment of the cathode circuit.

The final form of uhf oscillator we shall consider is the ring oscillator. This circuit was used early in radar development. Tubes of high peak-power capability had not been developed, and it was found impossible to obtain the desired high-power output (200 kw peak or above) from two tubes. Later, when tube development had progressed, the need for the ring oscillator largely disappeared. A typical form of the ring oscillator and its equivalent circuits are shown in Figs. 263 and 264.

The ring oscillator is essentially a combination of several push-pull oscillators and hence always employs an even number of tubes, usually 4, 8, or 16. The disadvantages are the multiplicity of tuned circuits, all of which must be adjusted, and the many possible paths for parasitic oscillations. The first problem may be mitigated by ganging the grid, plate, and cathode controls. The second is controlled by designing the circuit such that the parasitic paths are substantially longer than one quarter of the desired operating wavelength, thus producing lower frequency parasitics, and by arranging that the condition of oscillation is not attained at the lower frequency.

The advantage of the ring oscillator lies in the fact that several

small tubes, having low values of transit time and interelectrode capacitance, may be combined to achieve any desired peak-power output. In fact, since the tube capacitances are connected in

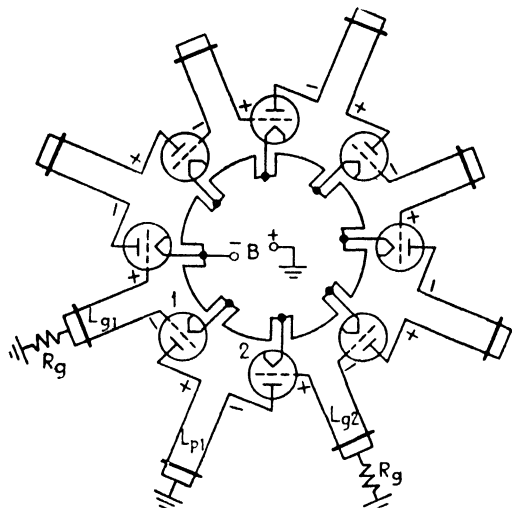



FIG. 263.—Eight-tube ring oscillator.

series, the upper frequency limit of a ring oscillator is higher than that of a two-tube push-pull oscillator using identical tubes and resonant lines. Ring oscillators have been designed to produce

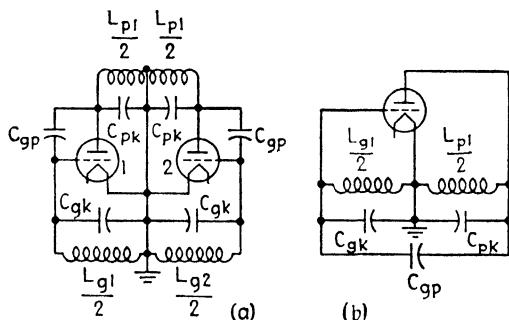



FIG. 264.—Two-tube (*a*) and one-tube (*b*) equivalent circuits of the ring oscillator shown in Fig. 263.

several hundred kilowatts peak power, at frequencies up to 400 megacycles.

Circuit analysis of the ring oscillator, as shown in Fig. 264, reveals that it can be reduced, like the two-tube oscillators

previously considered, to a one-tube equivalent circuit of the Hartley or Colpitts type. Hence the operating conditions previously enumerated for push-pull oscillators apply without change to each pair of tubes in the ring oscillator. The only additional condition of operation is that adjacent tubes in the ring must operate in opposite phase, that is, the potential of the grid of one tube must be increasing while that on the adjacent grid is decreasing. This is accomplished by alternating grid and plate connections, as shown in Fig. 263.

The output connection of uhf triode oscillators may be derived conductively, capacitively, or inductively. In the tuned-cathode two-tube oscillator, it is convenient to take the output from

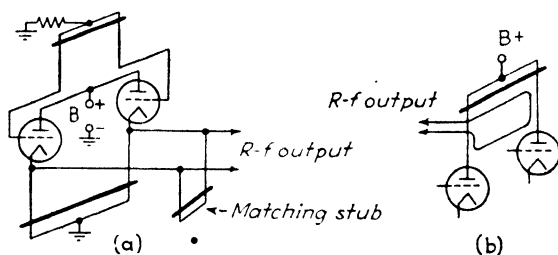



FIG. 265.—Methods of coupling to transmission lines.

the cathode lines, since they are at d-c ground potential. In the tuned-plate circuits, isolation of the direct voltage from the output transmission line is provided by a pickup loop. In the tuned-plate ring oscillator, the currents in all the plate-line shorting bars flow in the same direction, and a circle formed by the bars forms a virtual closed circuit to which the pickup loop may be coupled. These output coupling methods are illustrated in Fig. 265. The output transmission line is generally tuned by an adjustable stub to match the output impedance of the oscillator.

A problem arises when it is desired to employ a coaxial transmission line between oscillator and radiator. The push-pull and ring oscillators just considered provide inherently balanced-to-ground output. The conversion of the double-ended output to a single-ended coaxial line is performed in a balanced-to-unbalanced transformer ("bazooka" or "balun"). Two forms of these transformers are shown in Fig. 266. When conductive coupling is used, the inner conductor of a half-wave length of a

coaxial line is connected between the two resonant line conductors. This assures that whatever impedance is present at one conductor is transformed without change to the other, so that the resonant line remains balanced. The inner conductor of the output coaxial line may then be connected to either of the line conductors, and its sheath bonded to the sheath of the half-wave coaxial line. When loop coupling is employed, balanced-to-unbalanced transformation may be achieved by tuning the loop in series, as shown in Fig. 266b.

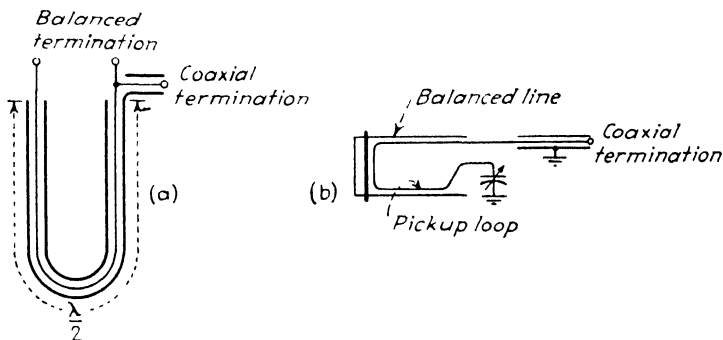



FIG. 266.—Transformations from balanced to coaxial feeds.

153. The Cavity Magnetron.—The principal transmitting oscillator used in the microwave frequencies (50 cm and shorter wavelength) is the cavity magnetron. This adaptation of the conventional magnetron was developed at the University of Birmingham by Oliphant and his coworkers. This type of magnetron contains a number of resonant cavities forming the anode, surrounding a centrally located equipotential cathode. An axial magnetic field permeates the space between anode and cathode, causing the emitted electrons to follow cycloidal paths in the space between the two electrodes. A majority of these electrons travels past the entrances of the resonant cavities and maintains the oscillations within them. By proper proportioning of the electric field (anode potential) and of the magnetic field with respect to the tube geometry, high efficiency (up to 70 per cent) may be achieved.

The construction of a typical cavity magnetron is illustrated in Fig. 267. The anode is a block of copper (either solid or built up of laminations) into which have been cut an even number (from 6 to 18) of cavities. Each cavity is connected with the central

chamber by a slot or aperture. The dimensions of each cavity and slot are such that their natural resonant period in the lowest mode is that of the desired output frequency.

Located within the anode is a cylindrical oxide-coated cathode, which is heated by a conventional heater wire. Attached to the anode is a series of wires called "straps," each one of which connects from one edge of one slot to the corresponding edge of an adjacent slot. The purpose of these straps is to reduce the number of modes at which the assembly of cavities can resonate simultaneously and thus to assure maximum power output at the desired frequency.

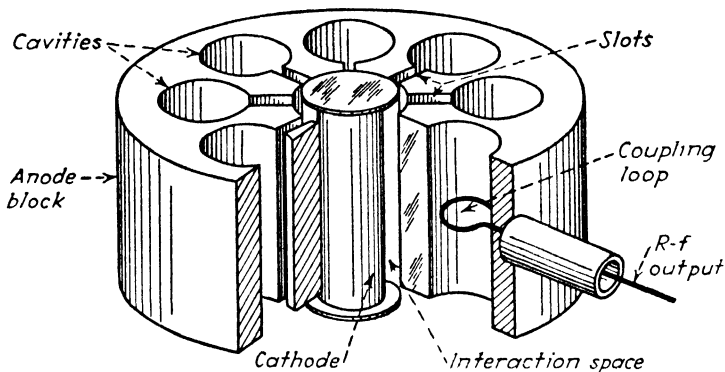



FIG. 267.—Internal structure of a multicavity magnetron.

Within one of the cavities is located a coupling loop which abstracts power from that cavity and, indirectly, from the other cavities which are effectively coupled to the first. The whole structure is enclosed in a metal envelope, exhausted to low pressure (a few hundredths of a micron of mercury) and placed between the pole pieces of a magnet, so oriented that a field of several thousand gauss passes through the anode-cathode space, parallel to the axis of the cathode.

In operation the anode is grounded (thus reducing the hazard of the anode potential) and a high negative voltage is applied to the cathode. If the applied potential has the proper value with respect to the magnetic field and the tube geometry, oscillations in the cavities are excited and sustained. Radio-frequency energy of the desired frequency then appears across the output coupling loop and its associated coaxial line.

The cavity magnetrons used in radar transmitters are designed

for pulse service, with duty cycles from 0.001 to 0.005 and with peak-power output as high as several million watts. Cavity magnetrons capable of producing c-w (continuous-wave) r-f energy are feasible, and can produce c-w power at levels as high as several kilowatts.

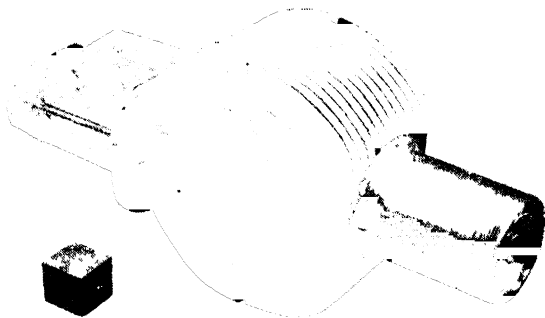



FIG. 268*a*.—Type 4J28 fixed-tuned magnetron, with coaxial output (*right*). This magnetron will generate 800-kw pulses at 1300 mc (see Table XI). The size is indicated by the 1-in. cube.

FIG. 268*b*.—Type 5J26 tunable magnetron, similar to the tube shown in Fig. 268*a* but tunable over the range of 1220 to 1350 mc. Note gear drive of tuning mechanism.

Cavity magnetrons have been designed to operate over the frequency range from about 400 megacycles to 26,000 megacycles. A particular magnetron is, of course, limited to a narrow range of tuning (about 1 per cent of the carrier) by the fixed nature of the resonant cavities. Groups of magnetrons, which have otherwise similar characteristics but which have slightly different

and over-lapping output frequencies, are produced to cover each of the frequency bands used in radar apparatus. Tunable magnetrons, in which the size or shape of the resonant cavities can be adjusted mechanically, cover a range of about 10 per cent of the design frequency. The external appearance of several cavity magnetrons is illustrated in Figs. 268 and 269.

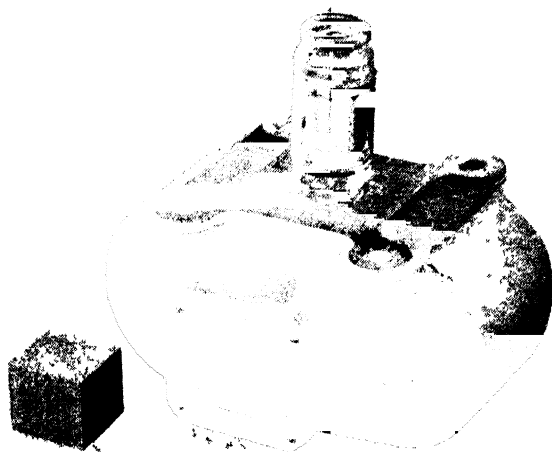



FIG. 269a.—Type 4J50 magnetron with integral magnet structure This tube operates at 9400 mc (fixed tuned) with a peak power output of 60 kw.

154. Internal Action of the Cavity Magnetron.—Many of the operating characteristics of the magnetron may be explained on the basis of a simple theory of its internal operation. Here we follow the notation of Fisk and his coworkers.[1] To simplify the geometry of the magnetron we first imagine that we have "unrolled" the cathode and anode from cylindrical form to two parallel planes, as illustrated in Fig. 270. The anode is shown, for the time being, without cavities. We inquire concerning the path taken by an electron emitted from the cathode and attracted to the anode by an accelerating field E volts per cm, which is normal to the electrodes, while passing through a magnetic field B gauss at right angles to the plane of the paper.

[1] FISK, J. B., and others. The magnetron as a generator of centimeter waves *Bell System Tech. J.*, **25**, 167 (April, 1946)

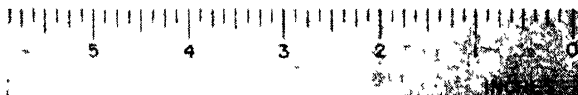
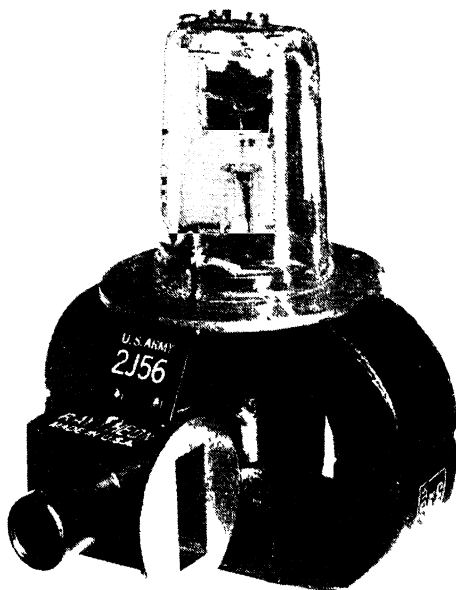



FIG. 269b.—Type 2J56 fixed-tuned 3-cm magnetron Note cylindrical shell of stub support which converts coaxial output to waveguide connector.

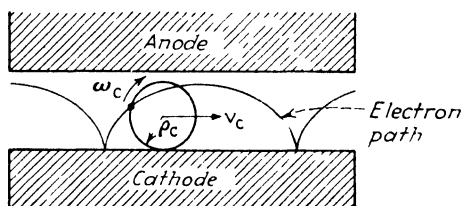



FIG. 270. Cycloidal motion of electron in crossed electric and magnetic fields

Under these circumstances, an electron emitted normally from the cathode describes a cycloidal motion, that is, it moves as if attached to a wheel which rolls without friction on the surface of the cathode. The radius ρ_c, angular velocity ω_c, and horizontal (axle) velocity v_c of this wheel are given by

$$\rho_c = \frac{mE}{eB^2} \tag{456}$$

$$\omega_c = \frac{eB}{m} \tag{457}$$

and

$$v_c = \frac{E}{B} \tag{458}$$

where e and m are the charge and mass of the electron. These relations describe the path pursued by the electrons under d-c conditions, that is, with unvarying fields.

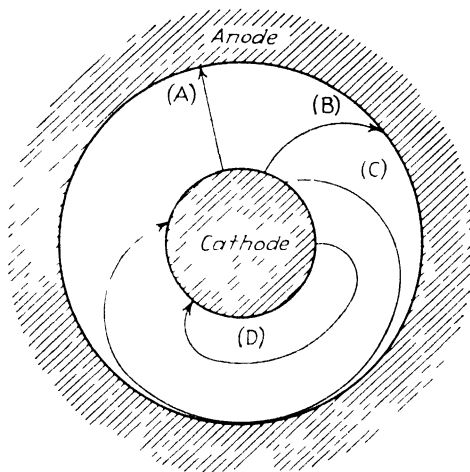



FIG. 271.—Electron paths between cylindrical cathode and anode, in order of increasing magnetic field (A) through (D)

Returning to the corresponding cylindrical geometry (Fig. 271), we note that the paths of the electrons vary markedly in accordance with the strength of the magnetic field. If, as in A, there is no magnetic field, the electrons pass directly from cathode to anode. A moderate value of magnetic field causes the path to curve slightly (as shown at B) but the number of electrons passing from cathode to anode per second is unchanged, so that the anode current remains constant. At a critical value of magnetic field, shown at C, the curvature of the electron path increases to the point where the electrons just graze the anode and are returned to the cathode. At this value of B, the current drops sharply. At higher values of B, no anode current flows at all,

because the electron paths are so constricted that no electrons reach the anode. In this latter case, shown at D, the electrons travel in closed loops, beginning and ending on the cathode surface.

To this d-c situation we must add the effect of r-f potential between adjacent segments of the anode, that is, across the cavity slots. The r-f electric field is disposed, as shown in Fig. 272, crosswise to the d-c electric field. When moving through both electric fields at once, the electrons tend to follow the rim of a

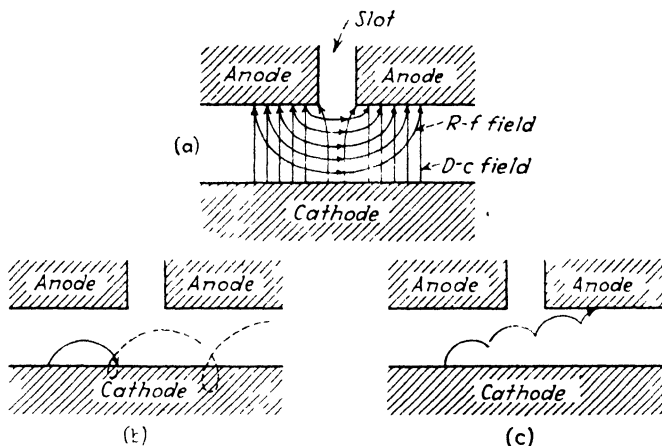



FIG. 272. Superposition of d-c and r-f fields (a), and resulting electron motions. In (b) the electron is accelerated and returns to the cathode. In (c) it is decelerated and gives up energy in passing to the anode.

wheel of increasing or decreasing radius. When the electrons are accelerated by the r-f potential, the loops in the electron path grow larger and the electrons are returned to the cathode where they are absorbed. But when the r-f potential retards the electrons, an opposite effect occurs, shown in Fig. 272b. Here the wheel has a decreasing radius and the electron pursues a wavy path between the electrode surfaces, losing velocity until eventually it is captured by the anode. This second type of electron path accounts for the generation of r-f energy, since the electrons gain energy from the d-c field and transfer the energy to the r-f field. In this manner the r-f oscillations are sustained within the cavities.

To examine this interaction between electrons and the r-f field in more detail, we reintroduce the cylindrical geometry (Fig. 273)

with either resonator slots in the anode. There are an infinite number of ways in which the r-f electric lines of force can be arranged between the successive slots, corresponding roughly to the infinite number of modes that may occur across the cross section in a cylindrical waveguide. The mode of greatest practical importance is the so-called "π-mode" shown in Fig. 273, in which the electric lines of force extend symmetrically from one anode segment to the next, the directions of the lines of force being opposite across adjacent slots.

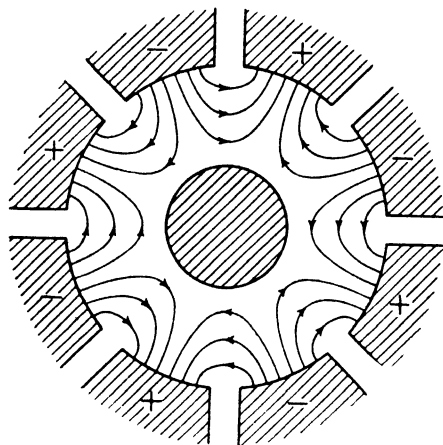



FIG. 273.—Distribution of r-f electric field across slots of an eight-cavity magnetron operating in the π-mode.

When these r-f electric lines are varying at radio frequency, a traveling wave of electric field appears to move about the center of tube at an angular velocity $d\theta/dt$ equal to the angular frequency of oscillation $2\pi f$ (where f is the radio frequency) divided by the number of resonator slots n.

$$\frac{d\theta}{dt} = \frac{2\pi f}{n} \qquad \text{radians per sec} \tag{459}$$

If we now imagine a cycloidal wheel rotating about the cylindrical cathode surface, the axle of the wheel travels about the center of the tube at an average angular velocity

$$\frac{d\theta'}{dt} = \frac{v_c}{r_a/2 + r_c/2} \tag{460}$$

where v_c is the axle velocity of Eq. (458) and $(r_a/2 + r_c/2)$ is the mean of the anode radius and the cathode radius. Oscillations

are sustained if the two angular velocities are the same, that is, if $d\theta/dt = d\theta'/dt$. Then

$$\frac{2\pi f}{n} = \frac{2v_c}{r_a + r_c} \tag{461}$$

But by Eq. (458) $v_c \cong E/B$ and E is approximately $V/(r_a - r_c)$, where V is the anode voltage. Substituting these relations,

$$\frac{2\pi f}{n} = \frac{2V}{B} \frac{1}{(r_a + r_c)(r_a - r_c)} \tag{462}$$

This equation may be rearranged to give the anode voltage required for oscillation in terms of the other quantities

$$V = \frac{\pi f r_a^2}{n} B \left[1 - \left(\frac{r_c}{r_a}\right)^2\right] \tag{463}$$

This derivation involves a number of approximations, particularly that which neglects the change in electric field with increasing radius. When account is taken of this and other pertinent factors the Hartree equation is obtained. This is like the above expression with an additional factor

$$V = \frac{\pi f r_a^2}{n} B \left[1 - \left(\frac{r_c}{r_a}\right)^2\right] - \frac{2m}{e}\left(\frac{\pi f r_a}{n}\right)^2 \tag{464}$$

This equation reveals a number of design and operating factors. We note that, if the magnetron has n cavities designed to resonate at frequency f and if the radii of anode and cathode are known, it is possible to plot a curve of corresponding values of V and B at which oscillation will occur. Such curves are known as "Hartree lines." For the π-mode we have been considering the lines are straight, and are found to be so in actual practice.

The Hartree equation does not give information regarding the current flow from cathode to anode. This quantity must be known, of course, before the efficiency of operation can be determined. The efficiency depends on the number of electrons which are retarded by the r-f field and hence give up energy to it, compared with the number which are accelerated and absorb energy. As we have seen, the retarded electrons are those which remain in the space between anode and cathode until they lose their energy and their velocity falls to the point at which they cannot escape capture by the anode. These captures constitute

the anode current. The electrons that are emitted in the wrong phase (that is, suffer acceleration) fall back to the cathode and do not contribute to the space current; but they do heat the cathode, having picked up energy from the d-c field, and this represents a power loss. It is important, then, in order to achieve high efficiency, that a maximum number of electrons shall not return to the cathode, but shall give up a substantial fraction of their energy before being collected by the anode. A very fortunate mechanism, known as phase focusing, assists in maintaining the escaped electrons in a region of maximum retarding potential Phase focusing is illustrated in Fig 274,

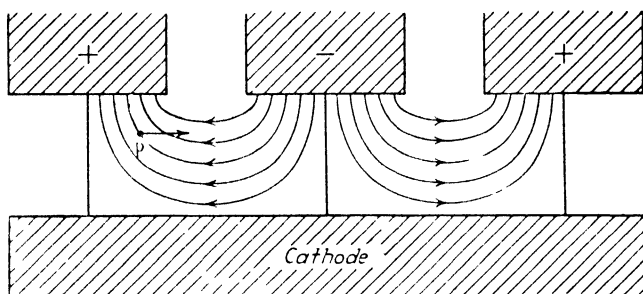



FIG. 274 Phase focusing

which represents two adjacent slots in the π-mode, which are plotted in the parallel-plane geometry The electron, at point p, is traveling from left to right against the lines of force and hence is being retarded. As it loses speed and tends to drift toward the anode, it encounters still closer spacing of the lines of force, that is, a higher retarding field, and is thus still further robbed of its energy before its eventual capture by the anode.

This brief analysis is based only on the π-mode that is used in most practical magnetrons. Many other modes of importance exist and may readily be excited simultaneously upon application of the anode voltage. This is particularly true if the magnetic field has too low a value, or if the impedance of the anode supply is not matched to the tube, so that an abnormally small anode current flows. Avoidance of such unwanted modes is an important practical problem that involves not only proper operating conditions in the magnetron but also the anode supply and r-f load. One important means of reducing unwanted modes is the insertion of conducting straps between alternate anode

segments. A full explanation of the theory of strapping is beyond the scope of this book, but it may be found in the literature.[1]

155. Operating Characteristics of Cavity Magnetrons.—The operating characteristics of the magnetron are conveniently studied in two groups: (1) those relating to the applied anode voltage and magnetic field, and (2) those relating to the connected load impedance. Typical characteristics of the first class are

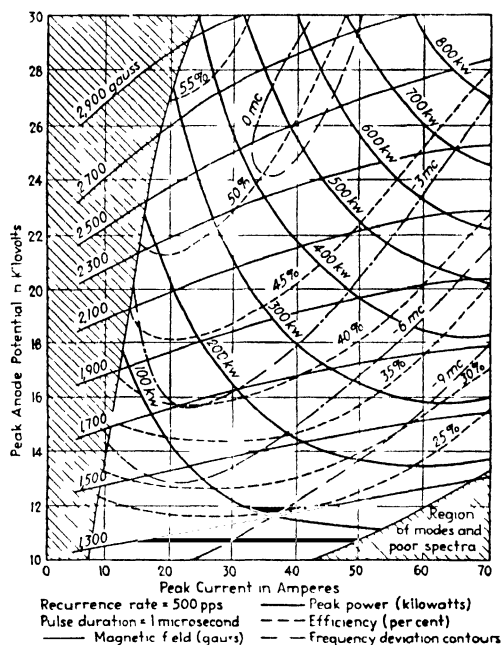



FIG. 275—Operating characteristics of a typical high-power 10-cm magnetron

shown in Fig. 275. Here the basic curve is a locus of corresponding values of applied anode voltage and resulting anode current, plotted for a given constant value of magnetic field.

Figure 275 shows that the stronger the magnetic field, the higher the anode voltage required to produce a given anode current. This is to be expected from Eq. (458), which states that the translational velocity v_c of the electrons in passing from cathode to anode is proportional to the applied anode voltage divided by the magnetic field. This velocity is a principal determining

[1] See reference, p. 381

factor in the anode current. The current, therefore, remains approximately constant as the anode voltage and magnetic field are increased simultaneously in the same proportion.

Contours of power output can be plotted conveniently on the current-voltage coordinates as shown in the figure. As might be expected, the power increases in the direction of the upper right-hand corner of the diagram, at correspondingly high values of anode voltage and current. Not so obvious is the fact that the efficiency increases in the direction of increasing magnetic field, tending toward the upper left-hand corner, as shown. The explanation of this effect is found in Eq. (456), which states that the radius of the cycloidal loops in the electron paths decreases as the magnetic field is increased. Thus, at high magnetic fields, the electron goes through a large number of loops in passing from cathode to anode. During each loop the efficiency of transferring energy from d-c field to r-f field is 100 per cent, since the electron begins and ends each loop at zero velocity. The loss of energy in falling to the anode during the last (partial) loop of the path is thus a small proportion of the whole energy transfer, when the magnetic field is high. This is another way of saying that the anode efficiency is high.

The numerical values shown in Fig. 275 are instructive. The magnetron represented is of the so-called "megawatt" variety, having a maximum peak output power in excess of one million watts. Operating anode potentials from 20 to 30 kv are used, producing anode currents, during the pulse, of 30 to 60 amp.

The magnetic field lies between 2,000 and 3,000 gauss. The range from 1,300 to 3,500 gauss is typical of 10-cm magnetrons. The higher frequency types, generating 3-cm and 1-cm wavelengths, use higher fields, from 3,000 to 8,000 gauss, in the interest of producing a large number of cycloidal loops within the restricted cathode-anode space of the smaller tube.

A peak power of 1,000,000 watts can be obtained from the type 720 tube with a field of about 2,900 gauss, peak anode voltage 27 kv, peak anode current 60 amp, and efficiency 65 per cent. Operation of this tube at anode currents under 10 amp, or at magnetic fields under 1,700 gauss, is not recommended because of the tendency of the cavities to resonate in multiple modes.

The slope of the anode voltage-current curve for constant

magnetic field is the input impedance presented to the modulating source. The value of this quantity (300 to 1,500 ohms in pulse-type magnetrons) must be taken into account in designing the modulator. One important effect of the high value of this input resistance is the necessity of feeding high voltage to the anode to obtain high power. High-voltage modulating pulses must be obtained from high-voltage tubes and capacitors, or they must be stepped up in a high-voltage pulse transformer.

Special attention must be paid to the cathode in magnetrons because the heat energy delivered to the cathode surface is in part derived from the anode supply. This action occurs when certain electrons, accelerated by the r-f field, fall back to the cathode surface, having picked up energy from the anode field. This bombardment heats the cathode surface in approximate proportion to the average power dissipation in the anode circuit. In high-power tubes this effect may do serious harm to the cathode unless precautions are taken. The cathode is customarily started at full heater voltage, but after the anode voltage is applied the heater current is reduced and in some cases turned off altogether, the cathode heat being supplied solely by the bombardment of the returning electrons. Excessive cathode heating, particularly that produced by operation at an excessive duty cycle, seriously shortens the life of the tube. In some experimental work, *water-cooled* cathodes have actually been used to permit operation at high average power!

Another set of operating characteristics, that of the 3-cm type 2J55 series, is shown in Fig. 276. This tube is of the "packaged" type, that is, it is furnished with a permanent magnet built integrally with the tube. Hence the characteristics are plotted for but one value of magnetic field, that of the associated magnet. The shaded region to the left of the 3-amp abscissa is the region of multiple moding, as indicated by the increased bandwidth of the emission in this region.

To avoid the inefficiency and instability associated with low anode current, it is desirable to bring the tube to its full operating anode current as rapidly as possible, to maintain this value throughout the pulse, and follow with similarly rapid decline to zero current. This is another way of saying that the waveform of the modulating current (and voltage) should be as nearly rectangular as possible. If the modulation is rectangular, the

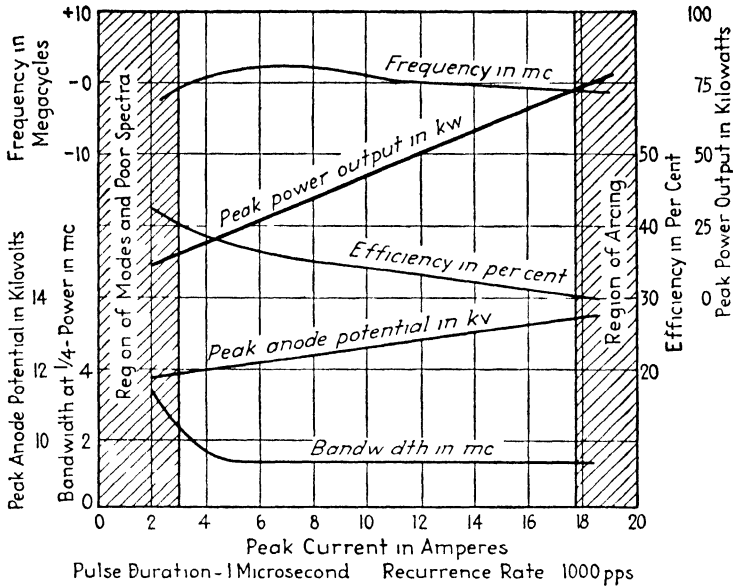



FIG 276 Operating characteristics of an integral-magnet 3 cm magnetron (type 2J55 series)

FIG. 277 —Typical magnet structures external type for *L*-band magnetron (*left*) and integral type

TABLE XI.—TYPICAL CAVITY MAGNETRON CHARACTERISTICS

Type no or series	Frequency range, mc	Maximum peak r-f power output, kw	Nominal peak anode rating	Maximum duty cycle	Maximum pulse duration μsec	Nominal average input power, watts
			L Band, 25–50 cm			
728 Ay-Gy	920–970 (fixed)	400	21.5 kv 35 amp	1/1,000	2	600
5J21-5J25	1,060–1,110 (fixed)	600	22 kv 50 amp	1/1,000	2	750
4J21-4J30	1,220–1,350 (fixed)	800	28 kv 60 amp	1/500	6	1,500
4J42	660–730 (tunable)	200	12 kv 10 amp	1/200	2	500
5J26	1,220–1,350 (tunable)	800	27.5 kv 60 amp	1/500	6	1,500
			S Band, 8–11 cm			
720 Ay-Ey	2,720–2,890 (fixed)	1,000	25 kv 70 amp	1/1,000	2	1,500
718 Ay-Ey	2,720–2,890 (fixed)	200	11 kv 22 amp	1/500	5	600
4J45-4J47	2,785–2,890 (fixed)	600	25 kv 45 amp	1/1,000	6.6	900
4J36-4J41	3,400–3,700 (fixed)	850	24 kv 43 amp	1/1,000	2.5	1,000
2J38-2J39	3,249–3,333 (fixed)	12.5	5.4 kv 5 amp	1/600	2	27
			X Band, 3 cm			
725 A	9,345–9,405 (fixed)	60	12 kv 12 amp	1/1,000	2.1	120
2J51	8,500–9,600 (tunable)	60	14 kv 10 amp	1/800	2	200
2J55-2J56	9,215–9,405 (fixed)	70	12.8 kv 12 amp	1/1,000	2.5	150
4J52	9,345–9,405 (fixed)	60	15 kv 20 amp	1/500	6	300

desired peak values of anode voltage and current are attained substantially instantaneously and the tube has no opportunity to oscillate in the unstable region.

As we have noted in Chap. II, the sideband energy outside the first zero is greater in a nearly rectangular modulating pulse than that of a rounded waveform, such as an error function or sine wave. Since energy outside the first zero of the spectrum does not contribute to the amplitude of the received echo pulse, a rectangular pulse wastes energy when, as usual, maximum received amplitude is the criterion. Moreover the outer sideband energy is a potential source of interference to other radars and other services. Up to the present, however, the stability achieved with rectangular modulation outweighs the disadvantages. In properly designed equipment, the magnetron attains full anode current amplitude in 0.1 μsec or less and falls to zero, at the conclusion of the pulse, in 0.4 μsec or less.

The magnetic requirements of the magnetron installation are simple. The recommended value of field should not vary over the face of the magnetron by more than about 5 per cent, and the field should, of course, be constant in time. These requirements are readily met by permanent magnets of the alnico type. The usual precautions, particularly avoidance of contact between magnet and ferromagnetic tools, etc., must be observed. Fig. 277 shows two typical magnet structures, one of the external type, the other an integral form.

Operating characteristics of typical cavity magnetrons are shown in Table XI [1]

156. The Rieke Diagram.—The operating characteristics just discussed are applicable when the magnetron is feeding a matched resistance load. When the output load impedance has some other form or value, the power output and output frequency are affected. A convenient representation of the effect of the connected impedance is the Rieke diagram, an example of which is shown in Fig. 278. This is essentially a circle diagram of the Smith type. The standing-wave ratio, on the output transmission line or waveguide, is plotted radially, and the position of the voltage minimum is plotted as an angle. On these coordi-

[1] FINK, D. G., Cavity Magnetrons, *Electronics*, **19** (1), 126 (January, 1946)

nates, contours of equal power output and frequency deviation are plotted as shown.

The center of the Rieke diagram, like the center of the Smith circle diagram, represents connection to a matched load. Any other point on the diagram represents a mismatch. Interpola-

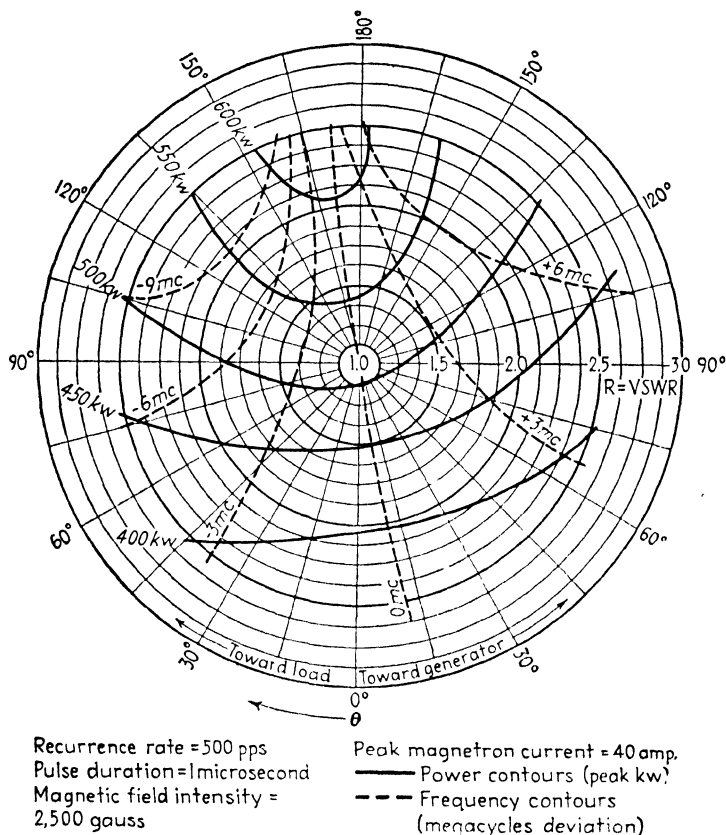



FIG. 278.—Typical Rieke diagram, showing power output and frequency a functions of the impedance of the connected load (type 4J46).

tion between contours at a given point reveals the power output and frequency for that particular load condition. The frequency and power output contours run approximately at right angles to each other, so that it is possible to vary the power output and frequency independently by adjustment of the load impedance. The frequency variation possible by this method is limited

to about 1 per cent of the design value (that at the center of the diagram).

The form of the frequency contours on the Rieke diagram can be derived from an equivalent circuit, in which the electrons are represented as an admittance Y_e driving a shunt-loaded tuned circuit, coupled through an ideal transformer to the load impedance, as shown in Fig. 279. When the corresponding values of load conductance and susceptance at constant frequency are plotted on rectangular coordinates, a straight line results. This line has the curved form shown in Fig. 278 on the coordinates of the Smith diagram.

157. Duplex Switching Systems (Transmit-receive Boxes).—The r-f circuit element that follows the r-f generator is the duplex

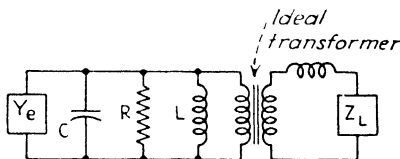



FIG. 279. Equivalent circuit of magnetron and load

switching system or transmit-receive (t-r) switch. The t-r switch permits the use of a single antenna for transmission and reception, at some loss in efficiency. The power loss incident to the use of a common antenna is commonly of the order of 2 to 3 db. This is generally tolerated in view of the mechanical and electrical convenience of a common antenna.

When a single antenna is used for transmission and reception, two precautions must be taken. First, the high power of the transmitted pulse must not be absorbed by the receiver circuit, not only because such absorption represents a loss of power, but more seriously because the absorbed power may do permanent damage to the receiver circuits. In the second place, the transmitter circuits must not absorb any appreciable fraction of the reflected echo signal, since such absorption would reduce the signal delivered to the receiver and would thus reduce the maximum detection range of the system.

This dual role of the t-r system indicates that a double-acting switch is required, that is, one which will open the receiver circuit while closing the transmitter circuit, and vice versa. A double-throw mechanical switch would serve this purpose were it not

for the fact that no mechanical switch can react in a fraction of a microsecond. Consequently it is necessary to employ electronic switching of the gas-discharge type.

The gas discharge has the property of extremely rapid action. Modern t-r tubes break down and become fully conducting within 0.01 μsec after receiving the excitation, and they recover after the pulse is removed in roughly 1 μsec. The disadvantage is the fact that the resistance of the path through the gas discharge is not negligible, as it would be in a direct-contact switch. Generally the discharge resistance is of the order of 10 to 50 ohms, depending on the nature of the discharge. When such a resistance is employed in coaxial line or waveguide circuits having

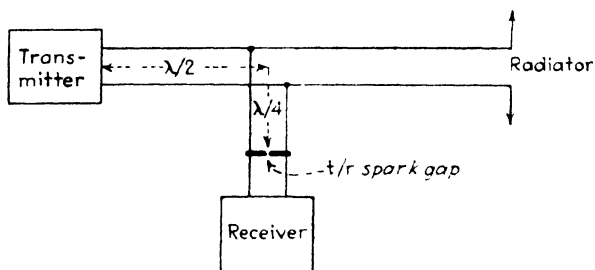



FIG. 280.—Basic duplex switch (t-r box).

characteristic impedances of comparable value, the short-circuiting action of the spark discharge is far from complete. This difficulty can be circumvented by the use of step-up and step-down transformers, in the form of quarter-wave line segments or cavity resonators.

The basic principle of the t-r box is well illustrated in a simple form used in the early l-f radars employing open-wire lines. Since the characteristic impedance of an open-wire line may be as high as several hundred ohms, the gap resistance of 50 ohms or less forms a substantial short circuit and no transformer is necessary. The typical circuit arrangement, shown in Fig. 280, is in the form of a T, consisting of a straight-through line from transmitter to radiator and of a sidearm to the receiver, placed one-half wavelength from the transmitter. In the sidearm, one-quarter wavelength from the junction, is placed an ordinary open spark gap, consisting typically of two rods bolted to the transmission line conductors and separated by about 1/50 in.

When the transmitted pulse is generated, sufficient energy is passed to the sidearm to cause the gap to break down (the required gradient in air at atmospheric pressure is of the order of 30,000 volts per in.).

The discharge constitutes a low resistance across the line, which has two effects. In the first place, the low resistance across the gap is transformed by the quarter-wave section to a much higher resistance at the junction of the T, so that very little power, beyond that required to maintain the spark, is delivered to the sidearm. In the second place, the power passed to the receiver is limited to V^2/Z_0 watts, where V is the voltage across the gap (generally not more than 50 volts after the gap has broken down) and Z_0 is the characteristic impedance of the line. Thus if the line impedance is 200 ohms and the gap voltage is 40 volts, about $(40)^2/200 = 8$ watts is passed to the receiver during the transmitted pulse, and this amount is independent of the power of the transmitted pulse. At low frequencies, when the input circuit of the receiver employs a vacuum tube r-f amplifier or frequency converter, this amount of power can usually be tolerated.

At the conclusion of the pulse, the excitation ceases, and the discharge ceases as a result of recombination of the gaseous ions (deionization). The time required for deionization is known as the recovery time, and is defined as the time, following the trailing edge of the pulse, at which the receiver power sensitivity has recovered half its normal value. In exposed air gaps of the type here discussed, the recovery time varies from 2 to 5 μsec.

Following deionization, the gap remains inactive and the echo signals subsequently received are conveyed to the T junction. Here the signal power divides equally between transmitter and receiver, if the transmitter impedance (then in the nonoperative condition) remains matched to the transmission line. This would be a serious loss of power Fortunately most transmitters display a marked change in impedance between the oscillating and quiescent conditions, so that the transmitter is not matched during the quiescent interval, and the major portion of the echo signal is conducted to the receiver. There is, however, always some absorption by the transmitter and this loss must be minimized. Proper choice of the length of transmission line between transmitter and T-junction will emphasize the transmitter mismatch. In the case shown this distance is taken as an even

number of quarter wavelengths, which is desirable when the quiescent impedance of the transmitter is higher than the oscillating impedance. If the reserve situation obtains, the distance should be an odd number of quarter wavelengths.

To minimize absorption during the quiescent interval, a second t-r switch, known as an "anti-t-r," is often employed. The circuit is shown in Fig. 281. During the quiescent interval, with the anti-t-r gap inactive, the short circuit beyond the anti-t-r gap is reflected through the half wavelength to the T junction

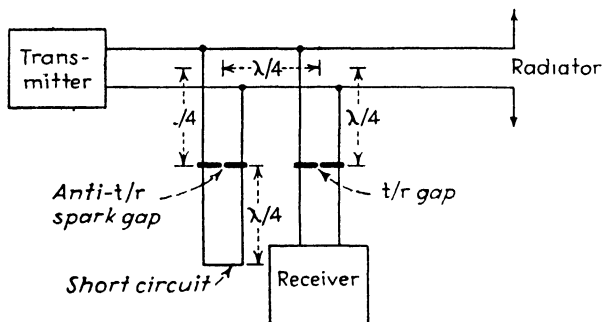



FIG. 281.—T-r and anti-t-r circuit.

where it appears as zero impedance. Consequently the transmitter output is shorted and very little echo signal power passes to the transmitter during the quiescent phase. During the pulse, however, both t-r gaps break down. The primary t-r gap, in the receiver sidearm, functions in the normal manner. The low resistance of the anti-t-r gap is transformed through the quarter-wave section to a high impedance at the junction with the straight-through arm of the circuit. Thus the short circuit on the transmitter output is removed, during the pulse, by the anti-t-r gap. An important practical advantage of this anti-t-r circuit is the freedom of choice in the length of line between transmitter and the anti-t-r box, which may have any value demanded by mechanical considerations.

The open-wire circuit just described has little but historical significance, because high-impedance open-wire lines are not used in modern equipment. When coaxial or waveguide transmission systems are used, even at low frequencies, the low characteristic impedance of the line or guide requires that a voltage step-up transformer be employed between line and gap

to transform the gap resistance to a value substantially lower than the characteristic impedance of the line. A simple transformer section for this purpose is illustrated in Fig. 282. It consists of a coaxial segment one-half wavelength long, short-circuited at each end, and terminated by two sidearms, one connected to the antenna and the other to the receiver. The short-circuited ends give rise to a standing-wave pattern within the coaxial chamber. The high impedance level is at the center of the chamber, where the spark gap is placed. When the gap breaks down, the magnitude of the current and voltage are sharply reduced, but their relative distribution remains unchanged.

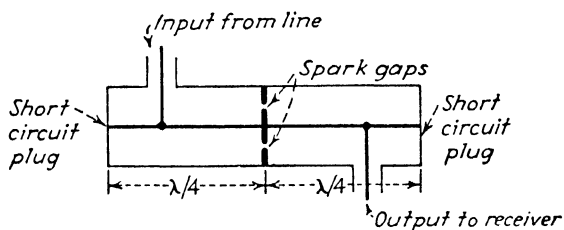



FIG. 282.—T-r transformer cavity for coaxial lines at low carrier frequencies (under 1000 mc).

Consequently, the impedance across the gap is transformed to a very low value, as viewed at the terminations of the chamber, and sufficient attenuation of the transmitted signal occurs to safeguard the receiver.

158. Transmit-receive Switches for Microwave Systems.[1]—As we shall see in Sec. 170, the most efficient system (that is, that with the lowest noise figure) for reception of frequencies of 1,000 megacycles and higher is the silicon crystal mixer. The rectifying action of the silicon occurs at interface between crystal contact. The contact interface may be damaged if the power fed to it during the transmitted pulse exceeds a certain limit. To assure reasonable life of the crystal unit power absorbed by the mixer should be not more than 200 mw. It is evident that no such simple t-r system as that discussed in the previous section is capable of this performance. Two additional steps are necessary. (1) An enclosed gap, with reduced gas pressure and consequently lower discharge voltage, is required. (2) A trans-

[1] SAMUEL, A. L., and others: Gas discharge transmit-receive switch, *Bell System Tech. J.*, **25**, 48 (January, 1946).

forming section is required to transform the gap resistance to a value small compared with the characteristic impedance of the connected lines. At microwave frequencies, the most convenient transforming section is a cavity resonator, fed at opposite edges at low-impedance points (through small coupling loops or slots). The gap is placed across the resonator in the region of highest voltage gradient. A typical microwave t-r switch of this type is shown in Fig. 283.

This type of t-r system makes use of a replaceable glass-enclosed gap, which, from its resemblance to more dignified electronic structures, is called a "t-r tube." Two widely used forms of t-r tubes are the type 721A used in 10-cm systems and the 724A used in 3-cm systems. The 721A tube, shown in Fig.

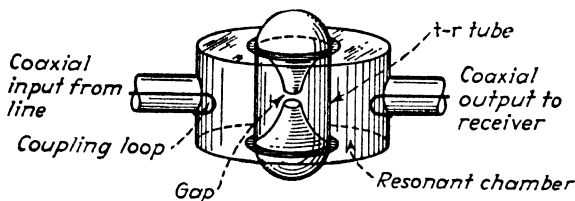



FIG. 283.—Typical microwave t-r tube and cavity.

284, comprises two conical electrodes which form the gap and an auxiliary electrode, known as the "keep-alive electrode," which is connected to a low-current, high-voltage d-c source and serves to maintain a supply of ions during the quiescent phase. The tube contains air and water vapor at a pressure of about 1 mm of mercury since these substances have the virtue of short deionization time (about 1 μsec in a new tube). In certain types, a radioactive salt is enclosed in the form of a coating on the back of one of the electrodes, to assist the keep-alive electrode in maintaining a copious supply of free ions between pulses.

A supply of free ions is essential if the gap is to react in the very short time (0.01 μsec or less) required to prevent excessive absorption of energy by the crystal mixer. Even when ions are present, a certain overvoltage appears across the gap before the discharge establishes itself. This "spike" in the waveform may exceed the normal by several hundred milliwatts. But if the spike is kept short, by maintaining sufficient ions just prior to breakdown, the energy (watt-seconds) absorbed by the crystal is kept within safe limits. Ordinarily the attenuation of the pulse

through a microwave t-r system should be at least 60 db (for example, from 100 kw to 100 mw).

Various t-r cavities are employed with the low-pressure t-r tubes. Figure 285 shows several typical examples used in coaxial and waveguide systems. The flanges of the t-r tube form integral parts of the top and bottom faces of the cavity. The electrodes form a central conductor, that is, the cavity is

FIG. 284.—Type 721A t-r tube for 10-cm systems.

of the coaxial type (Chap. III, Fig. 145). The cavity is tuned by a screw plunger set into the side wall, rather than by compression, since it is not practical to compress the glass envelope of the t-r tube.

Transmit-receive cavities for coaxial systems may be coupled either by coupling loops or slots, as shown in Fig. 285*a* and 285*b*, respectively. The smaller the loop or slot the higher the transformation ratio. The slot method of coupling is used in waveguide systems, as shown in Fig. 285*c*.

In systems requiring a high degree of t-r protection (as in systems of peak power above 200 kw) it is customary to employ an anti-t-r cavity, identical in function to the anti-t-r spark gap previously described. Figure 286 shows an anti-t-r system applied to a waveguide system. The distance between the t-r and anti-t-r junction points is one-quarter wavelength in guide. The anti-t-r cavity is identical to the primary t-r, except that it has no output termination.

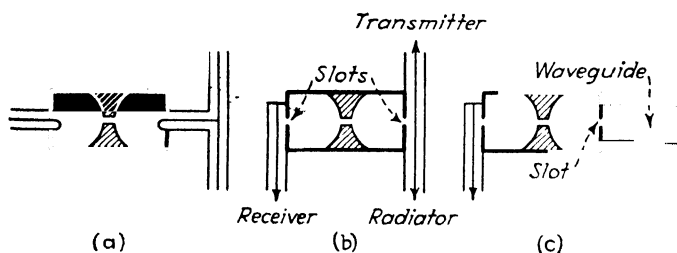



FIG. 285.—Methods of coupling to t-r cavities.

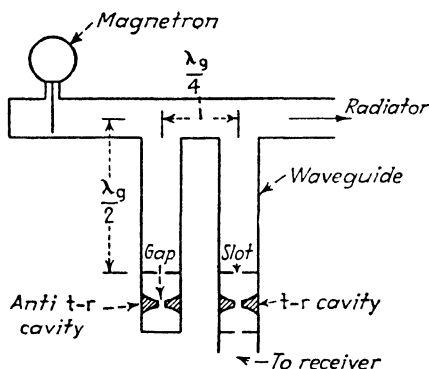



FIG. 286.—Waveguide t-r system, employing t-r and anti-t-r cavities.

159. Radiating Structures.—From the t-r switch the outward-bound signal travels to the radiator. The theory and design of directive radiators are discussed at length in Chap. IV. Here we shall discuss the effect of the radiator on the system to which it is connected, and describe practical radiators used in lobe switching, conical scanning, and high-speed scanning.

To insure maximum radiation of power, it is necessary that the impedance of the antenna be matched to the transmission line. For this purpose any of the tuning members described in Chap.

III may be used, such as tuning screws, apertures, adjustable stubs, T's, etc. The match is obtained by measurement of the standing-wave ratio, by the method outlined in Chap. XII. With care a voltage standing-wave ratio of 1.05 can often be obtained, so long as the radiator remains stationary.

When the radiator scans, if there are reflecting objects in its near vicinity, the impedance presented by the radiator to the transmission line varies. The effect on the standing-wave ratio is noticeable when the reflected signal is of appreciable magnitude (1 per cent or more of the radiated signal). Such reflections may arise, for example, if the housing ("radome") around the radiator is not placed symmetrically with respect to the origin of the scanning coordinates. For this reason, the radome should be made of material that is as nonreflecting as possible. If the radar beam encounters near-by massive objects (such as the superstructure of the ship, in the case of shipborne equipment) variations in the antenna impedance are inevitable.

These variations in radiator impedance are reflected through the transmission system to the transmitting oscillator and there cause the carrier frequency to shift slightly. The order of magnitude of the shift in the case of cavity magnetron transmitters may be obtained from the Rieke diagram (Fig. 278). This diagram shows that a standing-wave ratio of 1.5 may induce a frequency shift as great as 4 megacycles. When a standing-wave ratio as high as this is encountered, the receiver is thrown out of tune with the transmitter. Automatic frequency control of the receiver is a practical solution to this difficulty. The reflex klystron (Sec. 167), which is used as a local oscillator, may be tuned electrically over a range of 100 megacycles.

The presence of such large shifts in transmitter frequency indicates the need for a broad bandpass characteristic throughout the radiator and input circuits prior to the mixer. The broad-band stub-supported coaxial line is one suitable form; wave-guide systems usually possess sufficient bandwidth inherently. The antenna itself displays substantially constant impedance over an extended band when constructed of large conductors. The waveguide radiator is widely used as a paraboloid feed because it is not sensitive to frequency changes.

160. Lobe Switching.—One of the most important practical aspects of radiator design is the provision of overlapping lobes as a

means of refining the accuracy of angular indication. At frequencies under 500 megacycles, large mattress arrays are customarily used. Since it is not feasible to rotate such arrays about the target axis, as in conical scanning, means must be provided for obtaining two overlapping lobes and for switching from one to the other at a rapid rate. This technique is known as "lobe switching" (see Secs. 13 and 94, Chaps. I and IV, respectively).

The most direct means of providing two overlapping lobes is to employ two radiators, disposed at a slight angle to each other. Separate receivers are connected to each and the output of the two receivers is combined in opposed polarity in a single indicator

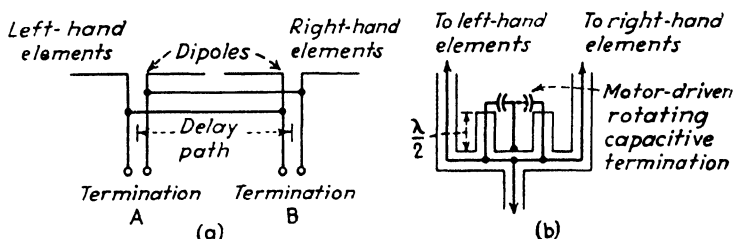



FIG. 287.—Methods of lobe switching.

that reveals the relative amplitudes of the signal as received on the two lobes. A better solution, technically, is to employ one radiator and one receiver, and to form the two lobes by sudden changes in the phasing of the elements of the array. Coincidentally with the phase changes, a corresponding displacement is introduced on the indicator screen, so that the signals from the two lobes are displayed side by side.

Figure 287 shows two methods of introducing the phase shift. In the first, two terminations are taken from opposite sides of the array (Fig. 287*a*). The radiator radiates and receives along a lobe at a slight angle from the normal to the array, since there is a phase delay incident to the passage of the signal from the array to the termination. This phase delay is from left to right in the case of the right-hand termination, and from right to left in the opposite termination. Hence the lobe looks off (squints) to the left for the left-hand termination and to the right for the right-hand termination. The two terminations are connected alternately to the receiver by an electronic switch

that operates synchronously with the displacement of the horizontal sweep of the indicator.

Figure 287*b* shows an alternative method when coaxial lines are used. Half-wave stubs with capacitive terminations are placed symmetrically in the feed lines to the two halves of the array. The capacitive terminations are varied alternately by a motor-driven rotor segment. The varying capacitance is transformed without change by the half-wave stub and causes successive and alternate delays in the two lines. The motor drive includes a cam contact that shifts the deflection of the indicator synchronously.

Lobe switching may be employed for another purpose: to fill in the gaps in the vertical coverage diagram. One system

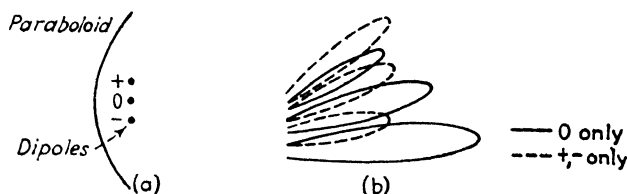



FIG. 288.—Vertical elevation of lobes to cover gaps in coverage.

of so doing is shown in Fig. 288. Here three dipoles and reflectors are used as the feed of a paraboloidal reflector. When a single dipole is excited, the beam is symmetrical about the axis of the paraboloid. By means of a solenoid-operated switch, the two outer dipoles may be connected in antiphase, that is, 180 deg in advance of and behind the center dipole, respectively. When this is done the beam is shifted in the upward direction by a slight angle (about 1.5 deg). This corresponds to half the angle between lobes, and thus in the antiphase position the lobes just cover the gaps of the in-phase position. When a target disappears into one of the gaps, the operator pushes the solenoid switch, shifts the lobes, and regains contact with the target.

161. Radiators for Conical Scanning.—In microwave radar, the small size of the radiating dipole makes the employment of conical scanning convenient (see Sec. 11) as a substitute for lobe switching. In conical scanning the axis of the radiated beam is separated by a small angle (about one-fifth the beam width) from the axis of the paraboloidal reflector. The beam is rotated about the reflector axis, tracing out a cone of such

dimensions that the lobe positions on opposite sides of the cone overlap each other.

The off-axis radiation is produced by displacing the feed (radiating source) laterally from the focus of the paraboloid.

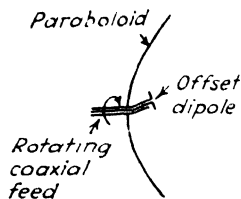



FIG. 289.—Offset radiator for conical scanning.

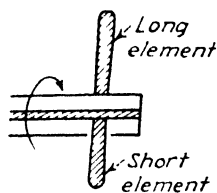



FIG. 290.—Asymmetrical dipole for off-focus feed.

The conical scan is formed by rotating the source about the focus as a center (or alternatively by rotating the paraboloid about the source). The displacement between source and focus can be introduced mechanically or electrically. A mechanically-displaced dipole and waveguide feed is illustrated in Fig. 289. It

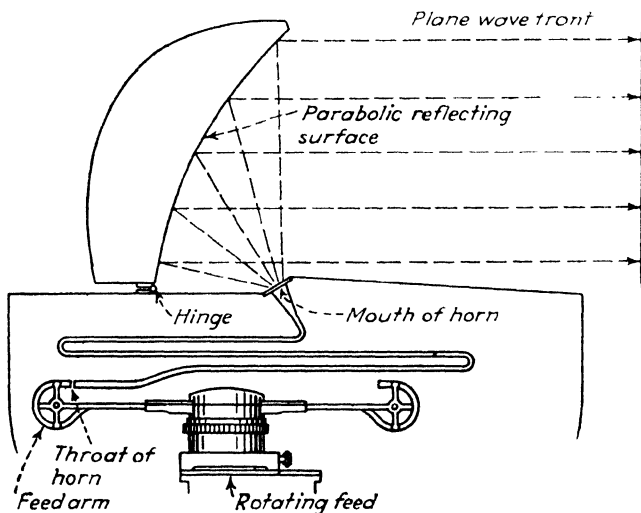



FIG. 291.—Rotating-feed type of high-speed scanner.

suffers from the vibration inherent in unbalanced rotating systems. A mechanically-balanced system is shown in Fig. 290. Here the displacement is effected by electrical dissymmetry. One half of the dipole is slightly longer than the other so that

the radiation arises effectively from a point off the axis of the paraboloid. The slight mechanical unbalance is readily compensated, and high speeds (up to 1,800 rpm) are readily attained with this type of rotating feed.

Nutating feeds are also used, particularly for spiral scanning (see Sec. 13) in systems requiring fixed polarization. The feed is of the waveguide type and is connected to the transmission system through a flexible joint. A motor-driven cam-operated linkage tilts the waveguide feed so that it nutates about the axis of the reflector. The plane of polarization is constant in this type of scanner, and thus fading and errors due to varying polarization are avoided.

162. High-speed Scanners.—When it is necessary to detect high-speed targets over a wide sector, ordinary methods of scanning are unsuited because they do not provide a sufficient number of observations per second to assure detection. For these purposes so-called "high-speed" or "rapid" scanners have been devised that will cover a section of up to 20 deg in less than $1/10$ sec. The sector is scanned completely at a rate higher than 10 per second and essentially continuous observation is provided. The cathode-ray tube has a medium-persistence phosphor (ordinary green screen).

Several high-speed scanners have been developed by Lewis, Foster, Wolff, and others, all of which employ the same basic principle but differ in mechanical form. The radiator is of the horn type, the mouth of which is elongated in the horizontal direction. The electrical field across the mouth of the horn is in effect a virtual current sheet. When the phase of the dipoles in this sheet is advanced linearly from one edge of the mouth to the other, the direction of the beam is shifted. As shown in Chap. IV, the beam direction is always normal to the equiphase surface; therefore the direction can be varied by adjusting the rate of change of the phase with distance across the horn mouth.

In practice, the change in phase is achieved by adjusting, in some mechanical manner, the distance from the primary source of radiation to the various points on the mouth of the horn. One simple system is the rotating feed used in the AN/MPG-1 radiator, shown in Fig. 292. Here the horn is folded back and forth upon itself to accommodate the necessarily small flare angle in confined space. The primary radiator comprises a

group of four waveguide radiators mounted like the spokes of a wheel and rotated past the throat of the horn. When one of the feed spokes is opposite the left end of the throat, it is nearest the left edge of the mouth and farthest from the right edge. Hence the phase of the virtual dipoles at the mouth varies from one edge to the other and the beam (normal to the equiphase surface) is displaced to the right. When the feed has rotated to the center of the throat, the equiphase plane coincides with the mouth and the beam is projected straight ahead. When the

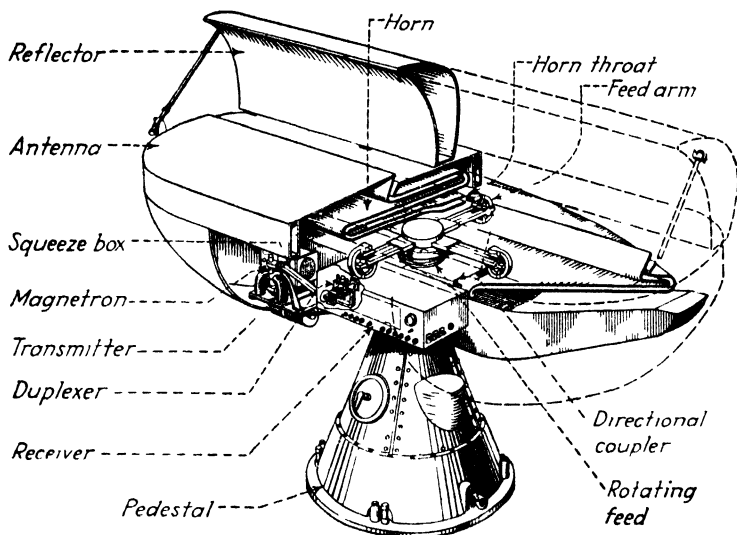



FIG. 292—Cutaway view of rotating-feed high-speed scanner (Type AN/MPG-1 radar)

feed reaches the right edge of the throat, the beam is displaced to the left. The feed then passes beyond the throat and is succeeded by another spoke that repeats the foregoing sequence. In this way the beam is swung over an angle of about 10 deg, disappears, and immediately reappears at the starting point. The spoke feed is rotated at 240 rpm, or 4 rps, so 16 spokes travel past the entrance mouth in 1 sec. The azimuth scanning rate is thus 16 scans per sec over a 10-deg sector, or 160 deg per sec.

A somewhat more complicated rapid scanner is the rolled-horn type developed by Foster and by Lewis. In this case the horn radiator is rolled up upon itself on a conical surface. Inside the inner surface of the roll is a smoothly fitting conical feed

member with a slit opening. This slit intersects the inner opening (throat) of the rolled horn, and as the feed member rotates, the intersection point travels from one end of the throat to the other. As a result, the phase of the virtual dipoles across the exit mouth varies linearly by an amount depending on the position of the feed member. With care in balancing the feed member, it may be rotated at 1,200 rpm, or 20 scans per sec, and the angle over which the beam moves may be as great as 20 deg. The rolled-horn rapid scanner is illustrated in Fig. 293. Rapid scanners have been used only in 3-cm and 1-cm systems, since their size is cumbersome at longer wavelengths.

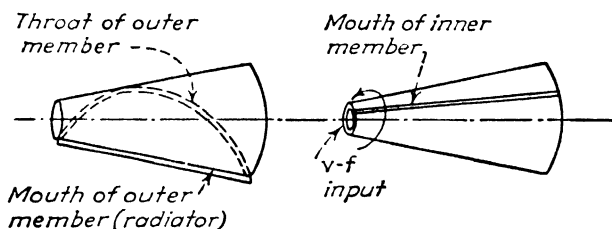



FIG. 293.—Rolled-horn type of high-speed scanner.

163. Nondirectional Radiators.—For certain purposes associated with radar nondirectional radiators are required. One such application is the radar beacon, a device used in conjunction with a radar to mark the position of, or otherwise identify, a particular point or target in the radar field of view. The beacon is an automatic receiver and transmitter that receives pulses ("interrogation") from a radar and returns a similar pulse or set of pulses ("response"). The beacon response may be on the same frequency as the radar; in this case the response is directly observable on the radar indicator. More usually the response is transmitted on a different frequency and an auxiliary receiver is provided at the radar. A particular form of radar beacon is the iff transponder ("identification, friend or foe"), carried by ships and aircraft to identify themselves to friendly radars. The response pulses are coded in accordance with a preset plan.

Evidently beacon receivers and transmitters must employ radiators that are nondirectional in the horizontal plane, since they must respond to interrogations arising from any quarter of the compass. Several forms of such radiators are widely used: vertical rods and vertical polarization, turnstiles, curved

dipoles, or loops for horizontal polarization. Figure 294 shows typical forms.

The vertical nondirectional radiator (coaxial and J-type quarter-wave rods) has a vertical coverage diagram in the form of a figure eight. In cases where vertical polarization is suitable, the vertical radiator is a simple structure, but its gain is low, of the order of 1.5 times in power relative to an isotropic radiator. When higher gain is required, several vertical dipoles may be stacked and fed in phase.

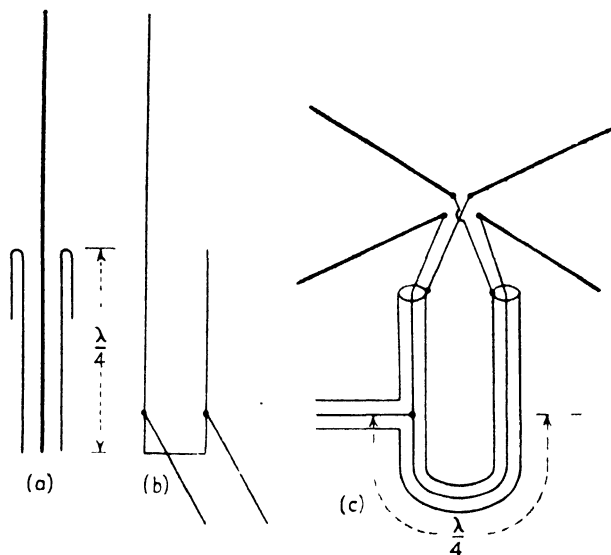



FIG. 294.—Radiators for producing nondirectional pattern in horizontal plane.

When horizontal polarization is desired (as when the interrogating radar employs horizontal polarization), stacked crossed-dipole radiators (turnstiles) are sometimes used. The two dipoles of each crossed element are fed 90 deg out of phase by suitable proportioning of the feed lines. Ease of adjustment is best served by the use of coaxial feed lines, including a curved member, one-quarter wave long, between the two dipoles. This member may be made adjustable (in the manner of the slide trombone) to secure quadrature phase between the dipole elements. The impedance match at the junction of the two feed lines is preserved by employing individual feed lines whose characteristic impedance is twice that of the main feed line.

The horizontal coverage diagram of a single turnstile element is the square root of the sum of the squares of two figure-eight patterns at right angles (Fig. 295), which is equivalent to a square with rounded corners as shown. The gain of a single turnstile element is about 1.5 and this value applies, in accordance with Fig. 295, in the horizontal direction. When higher gain is required, a number of turnstile elements are stacked, at halfwave separation.

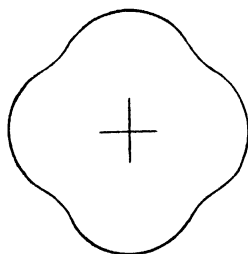



FIG. 295.—Horizontal pattern of crossed-dipole radiator.

A closer approach to the desired circular pattern may be achieved by three dipoles arranged in the form of a ring as shown in Fig. 296. Every other dipole element (quarter wave segment) is fed in phase from the center conductor, while the remaining elements are fed from the outer conductor.

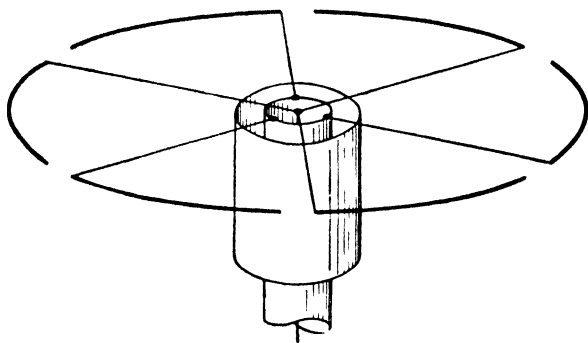



FIG. 296.—Triple-dipole feed with curved elements for nondirectional pattern.

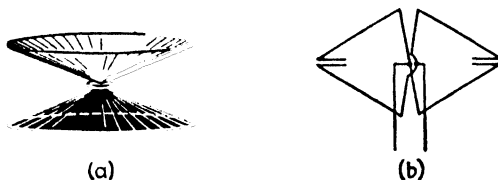



FIG. 297.—Biconical horn (*a*) and Alford loop (*b*) nondirectional radiators.

A quarter-wave choke section (bazooka) surrounds the outer conductor to prevent current flow along its outer surface, which would unbalance the dipole feed.

Other suitable forms of nondirectional horizontally polarized

radiators are the biconical horn and the Alford loop, illustrated in Fig. 297.

164. Radio-frequency Amplifiers.—After transmission and reception by the radiator, the radar signal returns via the transmission system and t-r switch to the receiver. The first r-f component in the receiver is the r-f amplifier. The use of r-f amplifiers is confined to radar receivers operating at carrier frequencies below 1,000 megacycles. At higher (microwave) frequencies the gain available from an r-f stage is so low that the stage introduces noise which would not be present if the stage were omitted. Successful r-f amplification at frequencies as high as 3,000 megacycles had been achieved in the laboratory at the end of the war but had not made its appearance in production equipment.

At lower frequencies, however, one or two r-f stages are generally employed. At the lower end of the radar band, 100 to 200 megacycles, the r-f stages are conventional negative-grid amplifiers, employing uhf pentode tubes, such as the acorn type. Above 200 megacycles light-house (disk-seal triode) tubes are customary, and the grounded-grid amplifier circuit is generally used.

The r-f amplifier must be carefully designed to minimize noise, since the noise generated in the stage may be the principal contribution to the noise figure of the receiver. Low-noise design implies obtaining the maximum possible gain in the stage and matching the input and output impedances to conserve power. The stage gain depends on the bandwidth passed; this must be minimized but there are definite limits below which the bandwidth cannot be set. First, the band must be wide enough (at least $2/d$ megacycles where d is the pulse width in microseconds) to cover the first zeros of the pulse spectrum. Secondly, additional bandwidth must be allowed for variations in transmitter frequency. This latter provision is particularly important in receivers employing automatic frequency control, since the frequency correction is confined to the local oscillator and i-f (intermediate-frequency) stages. To compensate for gross shifts in transmitter frequency the r-f stages must be tunable.

The design of a conventional negative grid r-f stage is illustrated in Fig. 298. Transformer coupling is usually employed

in the grid circuit in the interest of minimum grid-circuit noise. Autotransformer connections may be used, with the antenna line tapped above the grounded end of the grid coil. If a single-tuned circuit is used, elements of the equivalent tuned circuit are designed in accordance with the bandwidth equation given in Eq. (271), which may be rewritten as

$$Q = \frac{X_r}{R_{se}} = \frac{f_r}{2\,\Delta f} \qquad (465)$$

where f_r is the resonant frequency, $2\,\Delta f$ the desired bandwidth between 3 db points, X_r is the reactance of the inductive or capacitive element at resonance, and R_{se} is the series resistance.

FIG. 298.—Antenna coupling circuit.

In the interest of maximum gain, the shunt capacitance should comprise the tube input and wiring capacitances only. Tuning is accomplished by a variable inductive element, employing a movable metal slug.

As pointed out in Sec. 60, perceptible improvement (up to 3 db) in signal-to-noise ratio is possible by overcoupling the grid to the antenna circuit. This improvement is most readily obtained

FIG. 299.—Typical r-f amplifier for l-f radar use.

by empirical adjustment of the coupling while taking signal-to-noise readings.

The plate circuit of the r-f amplifier is designed on a similar basis, in accordance with Eq. (465). Here, however, we are concerned not only with thermal circuit noise but also with shot-effect noise from the electron stream with the amplifier tube. Shot-effect noise can be reduced somewhat by operating the tube so that dense space charge exists between the cathode

and grid. Adequate cathode emission and lower than average plate current contribute to this result.

Triode tubes are indicated for low-noise r-f amplifier service at high frequencies, because the interception noise is limited to one

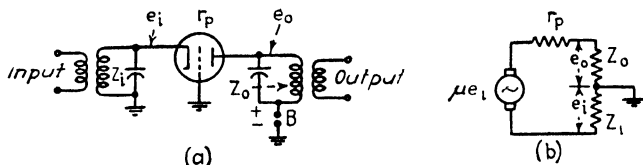



FIG. 300.—Grounded-grid circuit, actual (*a*) and equivalent (*b*)

grid, rather than two or three in the multigrid types. Figure 299 shows the details of a typical r-f stage for use at 200 megacycles, using the type 956 acorn tube.

The grounded-grid amplifier circuit is excellent for uhf amplification for a number of reasons. The actual and equivalent circuits are shown in Fig. 300. The input signal is applied to

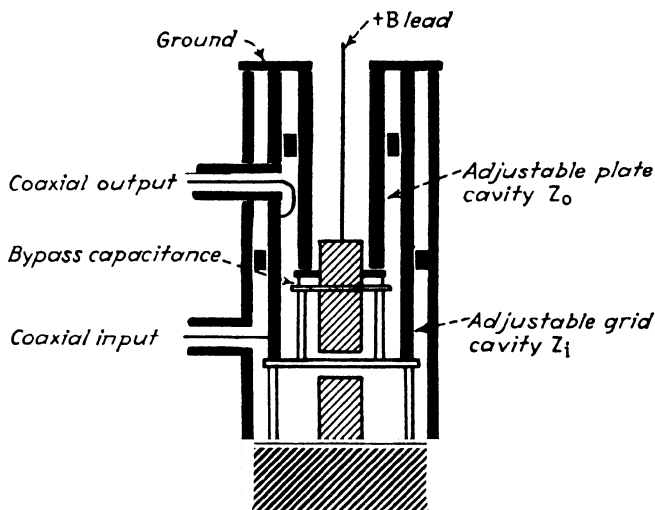



FIG. 301.—Cavities for disk-seal oscillator tube.

the cathode, which varies in potential relative to the grounded grid. This variation affects the space current in the normal fashion and the plate potential varies accordingly. No portion of the plate circuit signal is fed back to the grid so that the amplifier cannot oscillate. The grid-drive voltage is applied in

part, however, directly to the anode circuit without amplification, and this lowers the gain of the stage relative to that obtainable from the same tube in the conventional circuit at low frequencies. But at high frequencies the loss in gain is overbalanced by the isolation of plate and grid capacitances. This not only lowers the input capacitance, but prevents oscillation and makes neutralization unnecessary. The grounded-grid circuit is particularly adaptable to the disk-seal tube. A typical grounded-grid r-f amplifier using this tube is illustrated in Fig. 301.

Other specialized forms of amplifier tubes, notably the klystron and the inductive output tube, may also be used as r-f amplifiers. At frequencies above several hundred megacycles, the two-cavity klystron displays low efficiency as an amplifier, however, and the noise introduced is excessive. For this reason klystron r-f amplifiers, while operative at microwave frequencies, do not improve the noise figure of microwave radar receivers and are not used. The inductive output tube is used as an r-f amplifier at frequencies up to 600 megacycles with good results.

165. Superheterodyne Components.—Virtually all receivers used in radar are of the superheterodyne type (superregenerative receivers are used in certain radar beacons). The superheterodyne is effective because the major portion of the receiver gain is obtained at a frequency lower than the carrier value, and hence the amplification can be performed with simpler circuits, using fewer tubes and providing generally better noise performance than is possible at radio frequencies. A very important consideration in microwave receivers is the fact that the output noise from the silicon crystal mixer (superheterodyne converter) is less at intermediate frequencies than at carrier frequency, by some 6 to 10 db.

The usual components of a superheterodyne are employed: mixer, local oscillator, and i-f amplifier. The mixer operates with two signal inputs, the r-f signal (from the antenna or preceding r-f amplifier) and the local oscillator signal.

166. Local Oscillators.—The local oscillator in a radar receiver is basically the same as that in any superheterodyne receiver and it must meet the same general requirements. Unlike the transmitting oscillator, the receiver local oscillator must oscillate continuously so that its output may be combined with the echo signal regardless of the time at which the echo arrives. It

must generate an amount of power large compared with the noise power present at the input to the receiver, but the power must not be too great, lest it saturate the receiver or damage the crystal mixer. The frequency of oscillation must differ from the carrier frequency of the received signal, by an amount equal to the intermediate frequency (15 and 30 megacycles are typical i-f values in radar receivers). The oscillator must display sufficient frequency stability to avoid the necessity of retuning, and it should be susceptible of automatic-frequency control, to permit following casual variations in transmitter frequency. Finally the oscillator must be capable of being tuned over a sufficient range of frequency to follow gross changes in the transmitter frequency such as may occur when the transmitter is retuned to avoid interference or when the transmitter tubes are changed. Continuous tuning over a wide range (cf. the 3:1 range encompassed by a broadcast receiver) is not necessary, however.

The principal problem in the design of radar local oscillators is the attainment of the operating frequency. Power output of the order of 100 mw suffices. This amount of power can be generated in receiver tubes to which are attached appropriate forms of tuned circuits. At vhf (very high frequency) carrier frequencies (100 to 300 megacycles), lumped-constant tuned circuits may be employed in conjunction with vhf oscillator tubes (such as the acorn tube or door-knob tube), which have small electrode spacings and consequently small values of electron transit time. In the uhf range of frequency, lumped-constant circuits become impractical and it is usual to employ resonant cavities. Since conventional electron tubes are not easily connected to cavities, the disk-seal is used. This tube, already mentioned as a transmitting triode, has extremely small interelectrode spacing (0.003 in. between cathode and grid in the 2C40, for example) and is constructed with flanges that fit integrally into the resonant cavity faces. The upper limit of frequency at which tubes of this type can deliver stable power adequate for local oscillator service is about 3,000 megacycles.

In the microwave region, 3,000 to 30,000 megacycles, electron transit time difficulties become so great that it is necessary to resort to an electronic structure that separates the effects of transit time from the oscillating circuit. Two types of tubes

are suitable for this purpose, the magnetron and the klystron. In American radar practice, the klystron is used almost exclusively for local oscillator service at microwave frequencies. The type of klystron most widely used is the *reflex klystron.* The popularity of this tube arises from its susceptibility to electric control of operating frequency over a wide range, which suits it to automatic-frequency control systems. In view of the importance of the reflex klystron in this application, and because of its general applicability in shf techniques, we shall discuss its operation before considering specific forms of local oscillator circuits.

167. The Reflex Klystron.—The two basic forms of klystron are shown in Fig. 302. The double-resonator type (Fig. 302*a*)

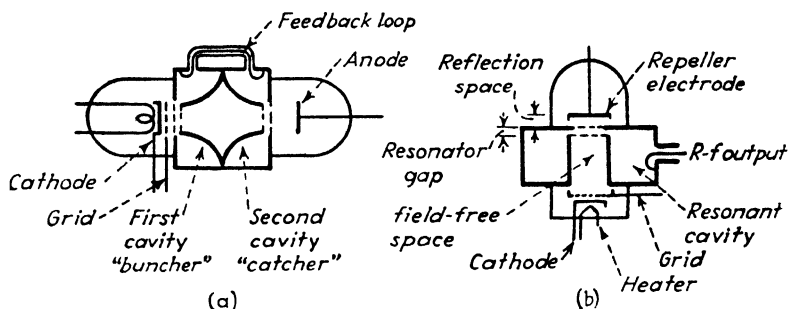



FIG. 302.—Klystrons: double-resonator type (*a*) and reflex type (*b*).

is used where high power is needed and electrical control of frequency is of secondary importance. The reflex form (Fig. 302*b*) achieves oscillation by passing the electron stream twice through the same resonator. The electrons are reversed by negative potential on a repeller electrode, changes in which cause corresponding variations in the oscillator frequency. Application of control voltage to the repeller electrode provides a simple automatic-frequency control. The same basic principles govern the operation of both types of tube. Since we are particularly concerned with the local oscillator application, we shall confine our attention to the reflex variety.

The cathode (Fig. 302*b*) supplies electrons which are focused by grid electrode, which also controls the value of beam current. The electrons in the beam are accelerated by the anode voltage until they pass the anode plane and enter the field free space within the central post of the resonator. The electrons in this

space maintain a constant velocity of

$$v_0 = \sqrt{\frac{2eE_0}{m}} \tag{466}$$

where e is the charge on the electron, m its mass, and E_0 the anode potential with respect to the cathode.

The electrons then enter the resonator gap, a narrow space between two grid structures that are parts of the resonator walls. The electric field in this gap is varying rapidly as a consequence of oscillations in the cavity. The variations in gap potential impose corresponding variations in the velocity of the electrons in the stream. It is assumed for purposes of this approximate analysis that the width of the resonator gap is very small compared with the longitudinal dimension of the electron path, and therefore the variation in velocity is instantaneously imparted to the electrons as they cross the gap.

The velocity of the electrons, on leaving the resonator gap, depends on the phase of the alternating voltage at the instant they cross the gap. Those electrons which cross when the alternating voltage is zero proceed with the initial velocity given by Eq. (466). Other electrons, crossing when the field has its maximum forward value (upper face positive) have the highest velocity on exit. Those entering when the field has its maximum backward value (upper face negative) leave with the least velocity.

The electron stream, leaving the gap with the continuous distribution of velocities, encounters the field of the repeller electrode acting against the positive field of the anode. The net potential acting on the electrons is $E_0 - E_r$, where E_r is the repeller potential. This potential acts on the electrons while they are in the reflection space, that is, while they travel a distance $2s_0$ from gap to repeller and back. The deceleration during this period is

$$a = \frac{e(E_0 - E_r)}{ms_0} \tag{467}$$

The total time the electrons remain in the reflection space is

$$T_0 = \frac{2v_0}{a} = \frac{2s_0\sqrt{2mE_0/e}}{E_0 - E_r} \tag{468}$$

During this time, the relative velocities imparted during the passage through the resonator gap are maintained. Those electrons which gained velocity in passing through the gap move forward toward those which passed through without changing velocity. Similarly those which lost velocity fall back. After a certain critical time, the electrons in the stream become separated into groups or "bunches," each bunch centered on electrons that passed through the gap when the gap voltage was zero. The electron stream thus becomes modulated in density as it traverses the reflection space.

On reentering the resonator gap, the electron bunches are in a position to give up part of their energy to the resonator and thus maintain the oscillations. The bunches impinge on the gap in synchronism with the cavity oscillations and if the phasing is proper, continuous oscillations are maintained. A maximum amount of energy is abstracted from the stream when each electron bunch reenters the gap at the instant the gap has its maximum backward potential (top grid positive), that is, when each bunch is decelerated as it crosses the gap.

The critical phase relationship is achieved when the total time T_0, Eq. (468), above, between leaving the gap and reentering it, is equivalent to $n - \frac{1}{4}$ cycles, where n is any whole number. The quarter cycle difference assures that the bunch, formed about electrons leaving at zero gap potential, will return when the gap potential is maximum and opposed to the motion of the bunch. The number of cycles occurring during the time T_0 is $T_0 f$, where f is the frequency of oscillation. Hence

$$n - \frac{1}{4} = fT_0 \tag{469}$$

and substituting Eq. (468)

$$n - \frac{1}{4} = \frac{2fs_0\sqrt{2mE_0/e}}{E_0 - E_r} \tag{470}$$

This is the basic equation of oscillation of the reflex klystron

It will be noted that there are an infinite number of conditions of oscillation, corresponding to discrete integral values of n. At each value of n, the frequency of oscillation is prescribed, not only by the physical dimension of the resonators, but also by the applied voltages E_0 and E_r.

In particular, if the repeller voltage E_r is changed, the frequency will shift in direct proportion to the quantity $E_0 - E_r$. If the shift is so great as to exceed the limits over which oscillations can be sustained, the oscillation ceases, but resumes when the shift in E_r introduces the next integral value of n. Thus we find a number of oscillation conditions dependent on the extent of the reflector space s_0, the resonant frequency of the cavity f, and the applied accelerating and decelerating voltage $E_0 - E_r$.

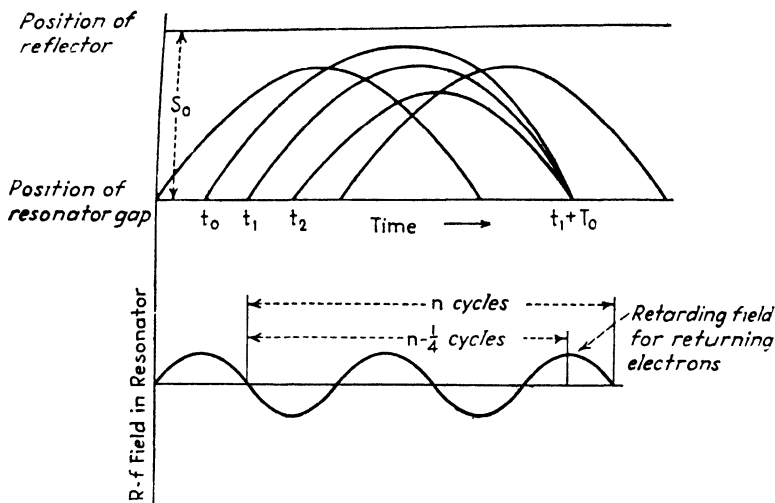



FIG. 303.—Position of electrons in reflector space at various points in r-f cycle. (*After Ginzton and Harrison.*)

Adjustment of the frequency can be made either by varying E_0 or E_r or both. Variation of the repeller voltage E_r is preferred, (1) because the current flowing in the repeller circuit is very small and hence the control circuit need supply little power, and (2) because the frequency sensitivity is greater to changes in E_r than E_0, since E_0 appears in both numerator and denominator in Eq. (470).

The factors governing the bunching of electrons and the transfer of energy from the bunches to the resonator are shown graphically in Fig. 303. At the bottom of the diagram is a sine wave representing the varying potential across the resonator gap. Electrons crossing the gap at time t_1, when the gap potential is zero, pass without change of velocity and return after a time T_0. Electrons crossing the gap earlier, at time t_0,

receive a loss in velocity. They remain in the reflection space a longer time and return to the gap coincidentally with those leaving at t_1. Still other electrons crossing at a later time, t_2, receive an increment in velocity, spend less time in the reflector space, and also return to the gap coincidentally with those leaving at t_0 and t_1. Hence the bunch is formed on return to the gap. If the time T_0, during which the t_1 electrons remain in the reflection space, is a whole number of oscillations less one quarter, the bunch is decelerated during its return crossing of the gap, and a maximum of energy is given up to the resonator.

The same principles govern the action of the double-resonator klystron (Fig. 302a). Here the distance between first and second resonator (the space within which the bunching occurs) is $d_0 = 2s_0$, and there is no deceleration between the two resonators. The condition for oscillation is

$$n - \frac{1}{4} = \frac{fd_0}{\sqrt{2eE_0/m}} \tag{471}$$

where the symbols have the same significance as in Eq. (470).

168. Conditions for Maximum Output in a Reflex Klystron.—Thus far we have examined only the mechanism of self-sustained oscillations in a klystron, and have stated simply that the output is a maximum when the bunch meets a maximum retarding field on returning to the resonator gap. To investigate more fully the condition for maximum output, we must compute the magnitude of the alternating current in the bunched beam, as it returns to the resonator gap. Let us consider an electron that crosses the gap initially (first transit) at a time t_1. If the voltage across the gap at that time is E_1, and if this voltage is small (as is usually the case) compared with E_0, the electron remains in the reflection space for a period equal to

$$T = T_0\left(1 + \frac{E_1}{2E_0}\sin 2\pi f t_1\right) \tag{472}$$

where T_0 is given in Eq. (468) and f is the frequency of oscillation of the resonator. The time t_2, at which the bunch returns to the gap is then $t_1 + T$ or

$$t_2 = t_1 + T_0\left(1 + \frac{E_1}{2E_0}\sin 2\pi f t_1\right) \tag{473}$$

Since the quantity of charge (equal to the current multiplied by the time interval during which the current flows) does not change in the beam, it follows that the outgoing steady current I_0 and the incoming bunched current I_2 have a ratio inversely proportional to the ratio of the time intervals during which the corresponding electrons passed the gap, that is

$$\frac{I_2}{I_0} = \frac{dt_1}{dt_2} \tag{474}$$

where dt_1 is the interval during which all electrons in the bunch originally passed the gap in the first transit, and dt_2 is the cor-

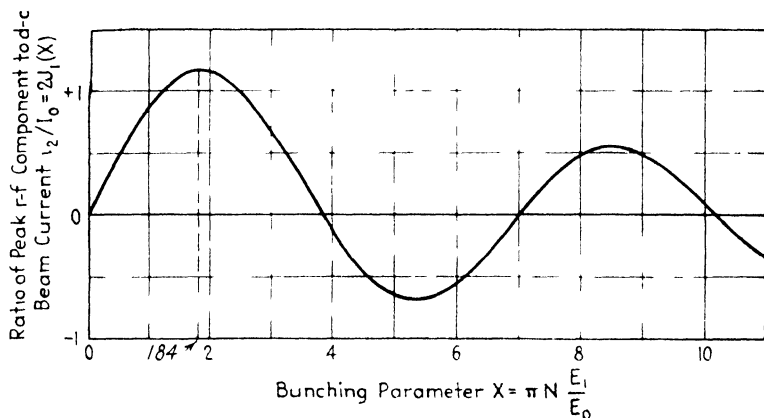



FIG. 304.—Relation of maximum r-f current to bunching parameter

responding time interval within which the bunch passed the gap in the second transit. The derivative dt_1/dt_2 can be obtained by differentiating Eq. (473) with the result

$$\frac{I_2}{I_0} = \frac{1}{1 + x \cos 2\pi f t_1} \tag{475}$$

where x is tle so-called "bunching coefficient," defined by

$$x = \frac{\pi N E_1}{E_0} \tag{476}$$

and N is the number of cycles of oscillations between times t_1 and t_2.

Equation (475) gives the peak value of the current in the bunched beam in terms of the average beam current I_0 and the

time t_1 at which the central electrons of the bunch left the gap on the first transit. This equation may be written in terms of t_2, the time at which the bunch returns to the gap, and this is a more pertinent relationship since it gives directly the sinusoidal variation in the bunched current. The relationship is

$$\frac{i_2}{I_0} = 2J_1(x) \sin (2\pi f t_2 - 2\pi N) \tag{477}$$

Here i_2 is the peak value of the r-f component of the bunched current, and $J_1(x)$ is the Bessel function of the first order and first kind, with the bunching coefficient as argument. A plot of the maximum value of i_2/I_0 against the bunching coefficient is shown in Fig. 304. It will be noted that the maximum bunched current, relative to the beam current, occurs when the bunching coefficient has the value

$$x = \frac{\pi N E_1}{E_0} = 1.84 \tag{478}$$

For maximum output, Eq. (478) must be satisfied, and this can be arranged, first by choosing values of N and E_0 to give a suitable value of E_1 and second by choosing the admittance of the resonator such that the resulting value of bunched current i_2 induces the same value of E_1 across the cavity. Further details on cavity design to insure this maximum relationship are given by Harrison and Ginzton.[1]

169. Practical Local Oscillator Circuits.—Figure 305 shows typical local oscillator circuits used in practice. The lumped-constant circuit in Fig. 305*a* is a modified Hartley circuit, with an acorn tube, which is useful at frequencies up to 200 megacycles. The light-house tube and associated cavities (Fig. 301) induce feedback from grid to plate cavity through two coupling loops joined externally. It operates in the uhf region (300 to 3,000 megacycles). The reflex klystron oscillator, Fig. 305*b*, requires no feedback loop, since the feedback is obtained by reversal of the electron stream. The power output is taken from the cavity by a coupling loop feeding a coaxial line.

[1] GINZTON, E. L., and A. E. HARRISON, Reflex Klystron Oscillators, *Proc. I.R.E.*, **34**, 97p (March, 1946). The notation used here is taken from this paper

The power generated in these local oscillators is generally 50 to 100 mw. Of this, only a small fraction is actually delivered to the frequency converter (mixer), since loose coupling between oscillator and mixer is highly desirable to avoid absorption

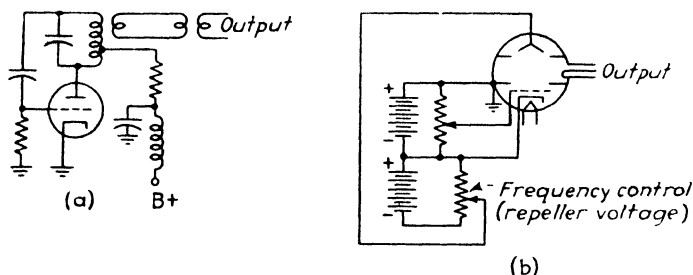



FIG. 305.—Local oscillator circuits: triode (a) and reflex klystron (b).

of the echo signal in the oscillator circuit. In microwave systems, about 2 mw is fed to the crystal mixer, a value well below the level (about 100 mw) that would injure the crystal interface.

Large changes in frequency are introduced mechanically by capacitive or inductive tuning in the lumped-constant circuits,

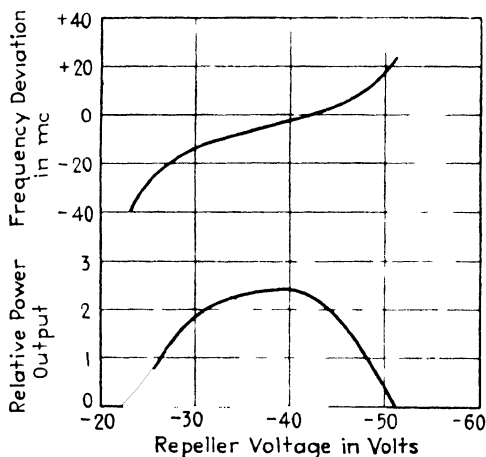



FIG. 306.—Effect of repeller voltage on frequency and power output of a reflex klystron.

or by tuning plugs or mechanical deformation in resonant cavities. Electrical tuning of reflex klystrons is illustrated in Fig. 306. The frequency sensitivity in 3,000-megacycles reflex oscillators is from 0.2 to 2 megacycles per volt change on the repeller

electrode. A tuning range as great as 100 megacycles may be obtained by electrical control. Great care must be taken to stabilize the repeller voltage to avoid unwanted frequency variations. Hum from the heater circuit may also be a source of frequency variations. Typical output and tuning characteristics of a reflex oscillator are shown in Fig. 306.

170. Heterodyne Frequency Converters (Mixers).— Perhaps the most critical aspect of radar receiver design is the heterodyne converter or mixer that generates the intermediate frequency from the carrier and local oscillator signals. The mixer is one of the most potent sources of noise in the radar system. Since every decibel of noise eliminated in this source is equivalent to a decibel increase in the transmitter peak power, great effort has been spent in reducing the noise figure of mixers. During the war years mixer noise was reduced some 15 db by successive improvements in design, equivalent to an increase in transmitter power of some 32 times without increasing the size, weight, or power requirements of the equipment.

The superheterodyne mixers employed in broadcast and shortwave receivers (pentode, pentagrid, triode-hexode) are not suited to the high carrier frequencies of radar because the noise of such converters is inherently high, due to the random interception of electrons by the multitude of grids, and because the conversion gain is lowered by high input capacitance and by transit-time loading effects. Pentodes having a high ratio of grid-plate transconductance to space current (such as the 954 and the 6AC7 pentodes) have low noise figures and have found some use in l-f radars, operating in the 100 to 200 megacycle range. But at higher frequencies, the diode form of converter has been used almost exclusively. Vacuum diodes have been used in some applications. The most successful form, employed almost exclusively in the microwave region, is the silicon crystal.

Aside from low noise figure, the principal requirements in mixer design are simplicity of tuning and appropriate by-passing of the r-f components to prevent their absorption in the i-f circuits. The basic equivalent circuit of a mixer is shown in Fig. 307. The input echo signal, from the t-r box, forms one input, and the local oscillator signal the other input. The output, suitably by-passed against radio frequency, transmits the i-f signal to the succeeding i-f amplifier stages. In l-f systems the mixer is

usually preceded by one or more r-f amplifier stages. The r-amplifier (Sec. 164) is indicated whenever its noise figure is substantially lower than that of the mixer, which is generally

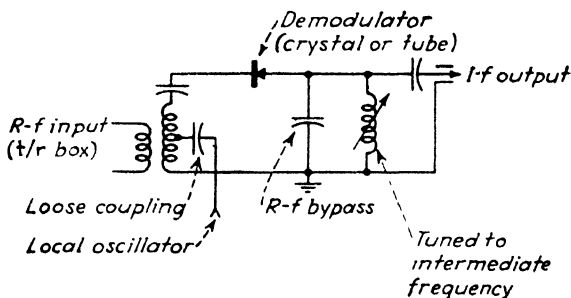



FIG. 307.—Basic crystal-frequency converter circuit (mixer)

true at frequencies below 500 megacycles, but becomes rapidly less so as the frequency is increased above this limit.

171. Silicon Crystals.[1]—The material used in crystal mixers is silicon. The crystal is mounted in a cartridge, the cross section of which is illustrated in Fig. 308. A small slab of silicon, containing boron or aluminum as an impurity, is mounted in the cartridge and a fine wire (catwhisker) of tungsten is mounted in contact with it. A contact pressure of several ounces exists between the metal and crystal surfaces. The contact point is sharpened to a fine point. The contact area is small (about 10^{-6} sq cm) and the capacitance between the two surfaces (which tends to by-pass the radio frequency) is thereby minimized. The thickness of the rectifying layer is a few microns, so small that transit-time effects are negligible.

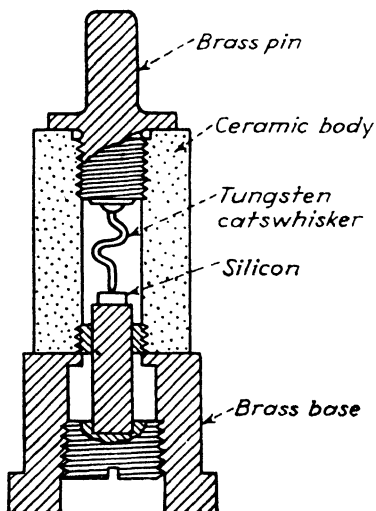



FIG. 308.—Structure of a typical silicon-crystal rectifier cartridge.

[1] STEVENS, W. E., Crystal Rectifiers, *Electronics*, **19** (7), 112 (July, 1946).

The electrical characteristic of a typical silicon crystal is shown in Fig. 309. This characteristic can be closely approximated by the expression

$$i = i_0(\epsilon^{\alpha V} - 1) \qquad (479)$$

FIG. 309.—Typical silicon-crystal rectifier current-voltage characteristic.

where i_0, the current at very large negative values of V, is of the order of 1 ma and the constant α varies from 2 to 20 per volt. The current-voltage characteristic varies with the temperature of the crystal.

The dissymetrical characteristic of the current flow is explained by the existence of a potential barrier at the boundary between silicon and the metal contact, which favors electron flow from metal to silicon. A similar barrier exists at the back face of the crystal and would nullify the rectification were it not for the fact that the back face of the crystal makes contact over an area several hundred thousand times greater than the area of the tungsten point contact.

FIG. 310.—Equivalent circuit of a crystal rectifier.

The equivalent circuit of the crystal rectifier (Fig. 310) consists of two resistances and a capacitance. The resistance and capacitance of the point contact form a shunt combination in series with the gross resistance of the silicon itself, the so-called "spreading resistance" through which the current flows to the point contact. At low frequencies, the capacitance (a few tenths of a micromicrofarad) may be neglected, but at microwave frequencies, the capacitance has an appreciable effect on the rectification efficiency. Typical values of efficiency are about 0.3 at 10 cm and 0.2 at 3 cm, when operated into an i-f load of 1,000 ohms.

From the design standpoint, the basic crystal specifications are

the noise contributed by the crystal and the amount of r-f power that may be safely applied to it. The noise contribution of the crystal arises from two sources: (1) the attenuation of the input signal as it passes through the crystal and (2) the thermal noise contributed by the crystal itself as a circuit element. The first effect arises from the fact that the conversion gain of a diode is always less than unity, that is, a loss. The conversion loss is defined as the ratio of the i-f signal available power at the output of the mixer to the available r-f power at the input. The loss is generally of the order of 5 to 8 db, or a power ratio of 3 to 6 times. The noise contributed by the crystal is expressed as a ratio of the noise actually present to the thermal noise of a pure resistance of the same magnitude. This ratio (sometimes known as the "noise temperature") varies from 1.5 to 4 times (from 1.7 to 6 db).

As indicated in Chap. X, the noise figure of a receiver is given by

$$n = 10 \log_{10}[L(R + n_{i\text{-}f} - 1)] \qquad \text{db} \tag{480}$$

where L is the conversion loss of the crystal, R the ratio of the crystal noise to equivalent resistance noise, and $n_{i\text{-}f}$ the ratio of the i-f amplifier noise to the equivalent resistance noise, all expressed as power ratios. At 3,000 megacycles, typical values are $L = 4$, $R = 3$, $n_{i\text{-}f} = 3$. Then n is $10 \log_{10} 20 = 13$ db.

These noise values apply at the intermediate frequencies generally employed in radar, 15, 30, or 60 megacycles. The noise is much higher, for a given bandwidth, if the intermediate frequency has a low value. For the same reason, poor noise performance occurs when crystals are used to rectify the input carrier directly to the video frequency region. It is thus advantageous to employ the superheterodyne principle with crystal rectifiers, and this is universal practice.

The allowable r-f signal power that may be applied to the crystal without injury is generally of the order of 100 mw. This corresponds to an input voltage of not more than a few volts and to rectified direct current not greater than 25 ma. Normal operating input power from the local oscillator is of the order of 1 mw, with the rectified current at 0.3 to 0.6 ma.

Higher values of power may be applied in the form of r-f pulses, without damage. But if the initial spike of power from the

transmitted pulse (that which passes through the t-r box before the gap breaks down) lasts for more than a few millimicroseconds, permanent damage may be done. Besides requiring protection from the transmitted pulse and local oscillator, mixer crystals

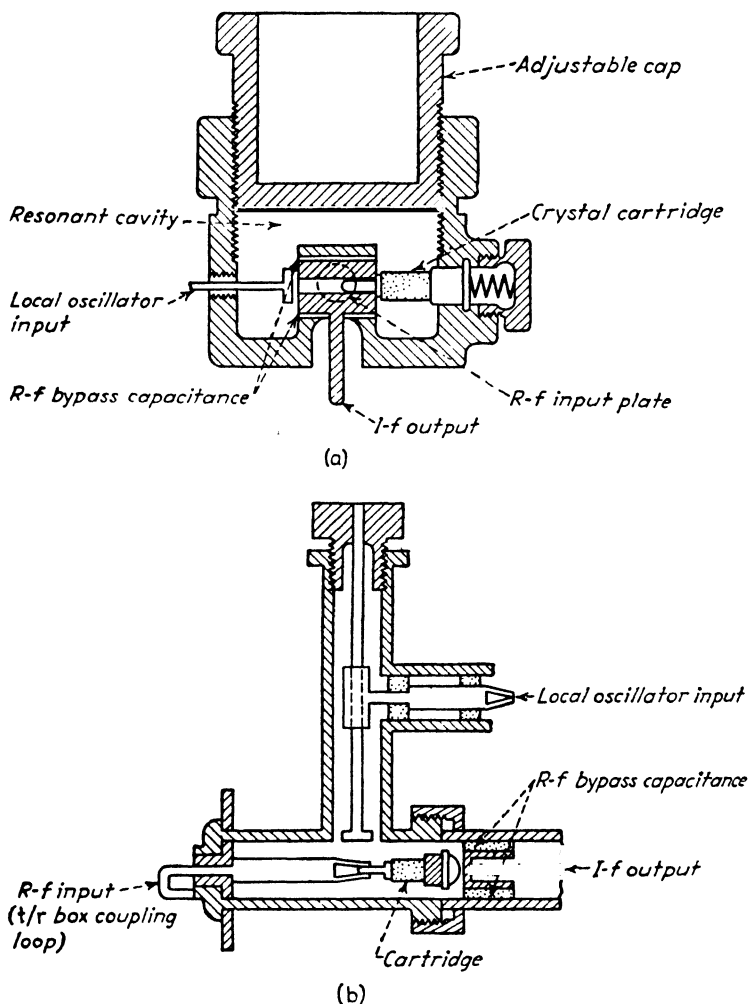



FIG. 311.—Typical microwave mixers: pot type (a) and pencil type (b).

must be carefully handled and stored. Mechanical shock, accidental discharge of static charge through the crystal during insertion in the crystal holder, and exposure to strong electric fields (r-f or video) are to be avoided.

172. Microwave Mixers.—Two forms of microwave mixers, designed for 3,000-megacycle systems, are shown in Fig. 311. The so-called "pot" mixer (Fig. 311*a*) is constructed as a coaxial resonant cavity, and has an adjustable screw top for tuning. The base of the crystal cartridge makes contact with the resonator wall, and the pin of the cartridge is engaged in a receptacle in the center post. The receptacle is insulated from the center post by a thin strip of mica, which by-passes the radio frequency. The i-f signal and direct crystal current are taken from an insulated lead connected to the receptacle. The signal carrier and local oscillator carriers are fed at opposite sides of the resonator through coaxial inputs that terminate in flattened probes adjacent to the center post. The r-f components flow between the center post and the resonator wall through the crystal.

The "pencil" mixer (Fig. 311*b*) contains the crystal cartridge within the coaxial i-f output lead, separated from it by a thin sheet of mica that acts as an r-f by-pass. The r-f input from the t-r box is conducted directly to the crystal through a coaxial segment and coupling loop. The local oscillator signal is inserted by a flattened probe adjacent to, but not touching, the signal input conductor. The closeness of coupling between local oscillator and mixer is controlled by the screw adjustment. No tuning is required, since the coaxial segments are of the broadband variety.

CHAPTER VIII

SYNCHRONIZATION EQUIPMENT

(Timers)

In the two preceding chapters we have examined, as separate items, circuits and structures widely used in radar systems. In this and the following chapters we examine the elements of the radar system in which these circuits and structures are integrated. For convenience these elements are divided into four groups: timers, transmitters and radiators, receivers, and indicators. The present chapter is concerned with the synchronization equipment (timers).

In early radar design, the timer was a discrete element of the system, occupying a separate chassis and considered apart from the elements operating under its control. In later designs, the timer has been absorbed in other elements of the system, particularly in the indicator circuits with which it is intimately related. In the latter case, the timing equipment may be considered as an auxiliary of the indicator. In other cases, the timing is performed by the rotation of a spark-gap modulator; in this case the timer may be considered an auxiliary of the modulation system. In all cases the functions of the timer are identical, regardless of the system used or its physical disposition.

173. Functions of Timing Equipment.—The basic function of the timing equipment is to coordinate the transmitter, receiver, and indicator. Specifically, the timer is called upon to perform some or all of the following (see block diagram, Fig. 312):

1. Produce a trigger that initiates the transmitter pulse
2. Initiate (or generate) the indicator deflection waveforms
3. Control the receiver gain by "gate" pulses
4. Produce intensifier pulses for brightening the indicator beam during the active portion of the sweep
5. Produce marker pulses for calibrating the range coordinate of the indicator
6. Provide a rigid, but adjustable, time relationship among the several pulses named above.

Underlying the function of time coordination is a basic timing source. Two types of timing are identified by the nature of this source: the *master-oscillator* system (externally synchronized) and the *self-synchronous* (internally synchronized) system. In the first, a master oscillator of the sine-wave, multivibrator, or blocking type is provided. This oscillator operates at low-power level, and is followed by successive amplifying and pulse-shaping stages. The function of the oscillator is to provide sine waves or pulses at regular intervals.

In the self-synchronous timing system, the basic timing is performed by the modulator (rotary spark gap) or the radio

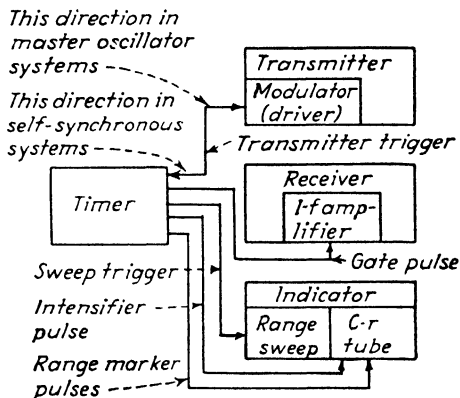



FIG. 312.—Function of the timer and associated pulses.

frequency oscillator (r-f blocking oscillator or squegging oscillator). The primary function is the generation of the transmitted pulse. A small portion of the energy of this pulse is used to control auxiliary circuits that coordinate the receiver and indicator with the transmitted pulse. In the latter system the transmitted pulse, or at least a portion of it, precedes the indicator sweeps and receiver gate. In the master oscillator system, on the other hand, differential delays of any required amount may be introduced between master oscillator pulses and the reaction of the transmitter, receiver, and indicator. In particular, the receiver gate and indicator sweeps may be initiated slightly in advance of the transmitted pulse.

The master-oscillator system is the more flexible of the two and is used where precise measurement of near-by targets is

required, as in gun-laying equipment. The self-synchronous system is much simpler from the circuit standpoint, and it is used in light-weight equipment, especially when high power must be combined with portability. The sequences of the timer pulses and waveforms in the two systems are compared in Fig. 313.

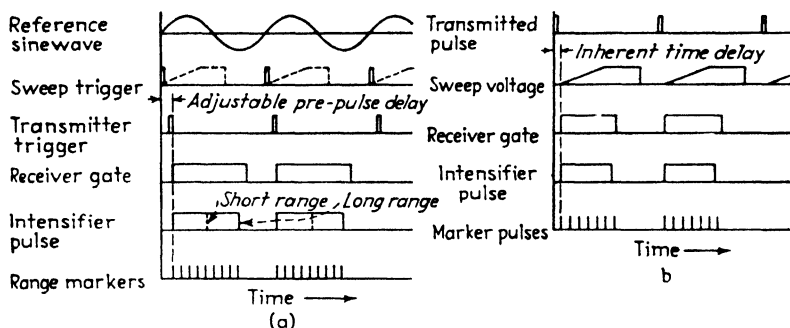



FIG. 313.—Typical waveforms of master-oscillator (*a*) and self-synchronous (*b*) timers.

174. Self-synchronous Timers. Rotary Spark Gaps.—The block diagram and simplified schematic of a typical self-synchronous timer are shown in Fig. 314. The basic timing element is a rotary spark gap, driven by a synchronous motor connected to the primary power source. The rotary gap

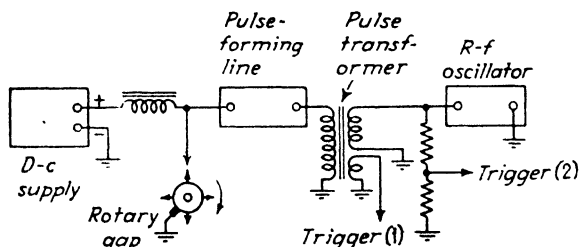



FIG. 314.—Typical self-synchronous timer (rotary-spark gap, type AN/TPS-3 radar). Alternative methods of deriving the trigger voltage are shown.

consists of one or more fixed electrodes and a number of moving electrodes spaced on the periphery of the rotor. When a rotor electrode and a stator electrode come sufficiently close, the high voltage between them breaks down the air in the gap. The current through the gap forms the modulating pulse. Coincidentally, a small portion of the modulating pulse is developed

across a voltage divider composed of resistor or capacitor elements. This smaller pulse acts as a trigger, which initiates the action of the indicator sweep generator, the intensifier pulse generator, the marker pulse generator, and the receiver gate generator. When a pulse transformer is employed between spark gap and r-f oscillator, an auxiliary winding is sometimes provided to develop the trigger voltage.

The essential specifications of the trigger are its repetition rate, rate of rise of the leading edge, and peak amplitude. The repetition rate is established by the number of junctures per second between stator and rotor electrodes. If there are n_s stationary electrodes, n_r rotary ones, and the rotor rotates at r rpm, the pulse rate is

$$f_p = \frac{rn_sn_r}{60} \qquad \text{cps} \tag{481}$$

In the equipment shown, the AN/TPS-3 early-warning radar, the spark gap has three rotor elements, one stationary element, and rotates at 4,000 rpm. The pulse rate is accordingly 200 pps. The spark gap is mounted directly on the shaft of the 400-cps primary power alternator.

In the rotary spark timer, the pulse-repetition rate is subject to two variations. First, the rate depends on the constancy of the primary alternating current that drives the synchronous motor. Slow variations of this sort are generally of no practical importance. The other variation occurs in the time of firing of each spark discharge. This time is subject to a variation of the order of 10 to 100 μsec, due to changes in the surface of the electrodes and the condition of the air between them. A supply of free ions must be available at the start of the discharge, and these may be provided by an auxiliary pointed electrode (trigger electrode or corona point) connected to the negative electrode. Even when this precaution is taken, timing variations of the order of 1 per cent of the pulse interval occur. These variations in the pulse interval would produce a blurring or "jitter" of the transmitted and echo pulses, as viewed on the type A indicator, unless each indicator sweep is initiated individually by the pulse whose echoes are displayed on it. In rotary-gap radars, therefore, the self-synchronous system of operation must be used.

Since the timing trigger is derived from the modulating pulse, the rate of rise, polarity, and amplitude of the trigger depend on the characteristics of the modulator. As described in Chap. IX, the spark-gap modulator employs an artificial transmission line to store energy between pulses. The shape of the pulse created by the discharge is determined by the magnitude and distribution of the inductance and capacitance elements of the line. Generally, the modulating pulse reaches full amplitude in a few tenths of a microsecond and the trigger pulse, across the voltage divider, has the same rate of rise. The polarity of the modulating pulse is usually negative (since cathode modulation is used in all magnetron transmitters and in a majority of the triode transmitters). The amplitude of the modulating pulse is generally 10,000 volts peak, or higher, and the amplitude of the trigger roughly one hundredth as great, depending on the amplitude required in the successive controlled circuits. The polarity may be reversed, if required, by passage through a pulse transformer.

One important result of the variation in pulse rate displayed by rotary-gap timers is the lack of synchronism between the received pulses and other radar pulses or jamming signals. Signals received from other sources appear blurred to the extent of the timing variations of the rotary gap and the associated indicator sweeps. This effect offers a means of distinguishing between true echoes and signals from other radars or from jammers.

Details of the construction and electrical characteristics of spark gaps and associated circuits for the modulating function, as distinguished from the timing function, are given in Chap. IX.

175. Self-synchronous Timers. Auxiliary Circuits.—The auxiliary circuits of the AN/TPS-3 self-synchronous timer are shown in Figs. 315 and 316. These are the type A indicator sweep, the intensifier generator, the ppi indicator sweep driver,[1] and the range marker generator. Since this radar is intended to be as compact and light in weight as possible, the circuits are as simple as their functions will allow.

[1] The ppi driver generates a single sequence of sawtooth waves. The division of the sawtooth waves into quadrature components and the establishment of the d-c reference are performed in other circuits that are, perhaps arbitrarily, considered a part of indicator system. These circuits are described in Chap. XI.

The operation of each of the auxiliary circuits is initiated by a biased ("one-shot") multivibrator. Refer first to the type-A indicator sweep and intensifier circuits, Fig. 315. The input trigger from the rotary-gap timer, amplitude about 80 volts negative with respect to ground, is applied to the left-hand triode of the multivibrator, which is normally conducting. The negative trigger cuts off this triode, and the multivibrator reacts, producing a rectangular wave of 100 volts negative amplitude in the plate circuit of the normally cut-off (right-hand) triode.

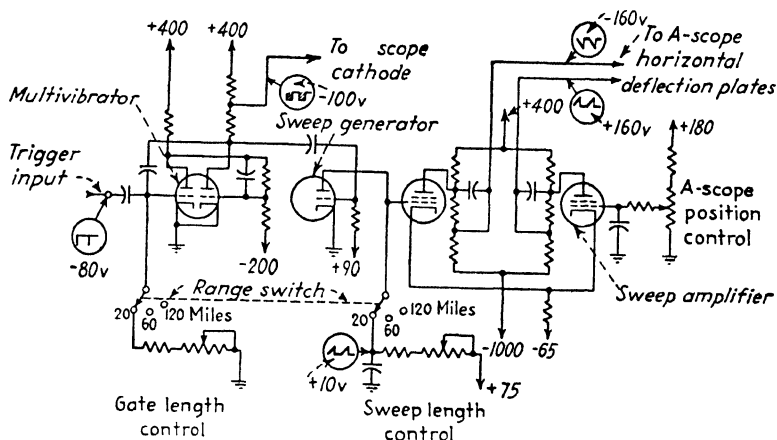



FIG. 315.—Auxiliary timer circuits of the AN/TPS-3 radar.

This negative rectangular wave, applied to the cathode of the type A indicator tube, causes the c-r beam to be intensified, immediately following the appearance of the trigger pulse, until the sweep is completed. Three lengths of sweep are provided, corresponding to maximum ranges of 20, 60, and 120 miles. A range selector switch inserts resistors in the grid of the normally conducting (left-hand) triode that have the proper value to generate negative waves of 214, 642, and 1,284 μsec duration. These are the echo intervals corresponding to the maximum ranges just mentioned. After the negative half waves are generated, the multivibrator remains inactive (due to the heavy negative bias on the right-hand triode) until the arrival of the next trigger. The time of response of the multivibrator is less than 1 μsec, so the type A indicator beam is turned on before the transmitted pulse (width 1.5 μsec) is fully completed.

The negative waves from the multivibrator also control the generation of sawtooth waves of lengths corresponding to the 20, 60, and 120 mile ranges. For this purpose, the negative wave is applied to the grid of a sweep generator triode, to the plate of which a capacitor is connected. The voltage across this capacitor is normally at a low value since the sweep generator tube is normally conducting. When the sweep generator tube is cut off by the multivibrator, coincidentally with the trigger pulse, the effective short circuit of the tube is removed and the capacitor voltage increases exponentially toward an upper limit of 75 volts positive. The initial portion of this exponential increase is sufficiently linear for the indicator sweep.

When the multivibrator completes the negative half wave, the sweep generator again becomes conducting and the voltage across the sweep capacitor is suddenly reduced to its initial value. Accordingly, there appears across the capacitor a sawtooth wave consisting of approximately linear voltage increases, of a duration equal to the echo interval for the maximum range then in use, followed by inactive periods, the whole sequence recurring 200 times per sec in synchronism with the trigger pulses.

The sawtooth wave across the sweep capacitor is passed through a push-pull (paraphase) amplifier that operates by virtue of a common unby-passed cathode resistor. Across the plates of the push-pull tetrodes the sawtooth waves appear in opposite polarity and 160 volts peak amplitude. These voltages are applied to the horizontal deflection plates of the type A indicator tube, causing the beam to deflect from left to right in synchronism with the trigger. Adjustable bias on the grid of one of the push-pull tubes regulates the relative amplitudes of the two sawtooth waves and thus permits centering the deflection on the screen.

The sweep and intensifying circuits of the ppi driver are almost identical to those of the type A indicator except for the magnitudes of the voltages. The sawtooth wave appearing across the sweep capacitor is fed, in this case, to a single-ended stage with shunt-fed plate supply. This is beam power stage which reverses the sawtooth waves and provides a heavy current output for driving the ppi rotary transformer (see Sec. 205, Chap. XI).

The remaining auxiliary circuits (Fig. 316) generate marker pulses to calibrate the range scales of the type A and ppi indicators. Here, again, the initial circuit is a biased multivibrator that produces a negative rectangular half wave of minus 70 volts peak amplitude and about 1,300 μsec duration. This wave is passed to the grid of a normally conducting triode, to the cathode of which is connected a tuned circuit of capacitance and inductance, tuned to 93 kc. This triode is suddenly cut off,

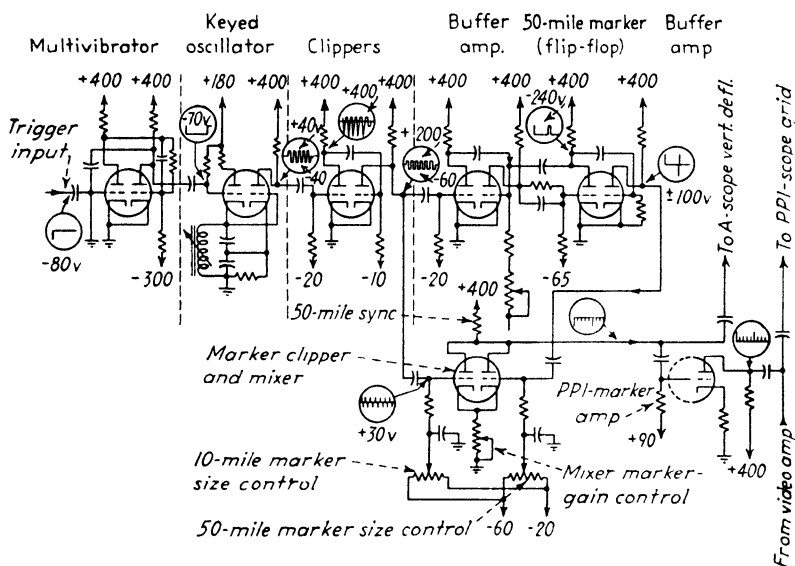



FIG. 316.—Range-marker generator circuits of AN/TPS-3 radar.

coincidentally with the trigger pulse, and oscillations are set up in the tuned circuit by shock excitation. The period of these oscillations (one full cycle) is 107 μsec, equal to the echo interval from a target at 10 miles. The oscillations tend to decrease in amplitude after the initial excitation, but remain substantially constant during the maximum echo interval (12 cycles, 1,284 μsec, 120 miles). The train of oscillations is first amplified and then passed to a clipper stage that produces rectangular waves from the sinusoidal drive. The rectangular waves are passed through a differentiating network, which develops short positive and negative pulses from the leading and trailing edges of the rectangular waves. The positive pulses, spaced at intervals of 107 μsec, serve as 10-mile marker pulses.

To this sequence of 10-mile markers, additional markers of somewhat greater amplitude are added at intervals of 535 μsec, representing range intervals of 50 miles. The 50-mile markers are generated by passing the rectangular wave from the clipper through a buffer amplifier to a biased multivibrator that acts as a 5:1 frequency divider. Pulses spaced at 535 μsec are developed from the output by passage through a differentiating stage. The 50-mile markers are combined with the 10-mile markers in a mixer stage consisting of two triodes with a common plate impedance. The combined marker signal is applied to the vertical plates of the type A indicator. After passage through a polarity reversing stage, the same pulse sequence is applied in positive polarity to the control grid of the ppi indicator tube, where it creates concentric range circles, the 50-mile markers being somewhat more intense than the 10-mile marks.

No provision is made in this timer for reducing the receiver gain during the transmitted pulse. Had a receiver gate been desired, however, it could be produced by shaping the trigger pulse to a length of about 5 μsec, in negative polarity. This pulse, applied to the grid of one or more of the i-f stages, would reduce the gain of the receiver within less than a microsecond of the appearance of the transmitted pulse. Alternatively, the trigger could be used to initiate the formation of a long positive rectangular wave (formed in a multivibrator). The i-f amplifier grids would then be operated normally cut off by a negative bias that prohibits reception during the transmitted pulse. The positive receiver gate, applied to the i-f amplifier grids after the transmitted pulse, would restore the receiver gain during the ensuing echo interval.

176. Master-oscillator Timers. Sine-wave Oscillator Operating at Pulse Frequency.—As a second example of timer equipment, we consider the synchronization circuits of the type SCR-720 radar, an airborne equipment used for short-range interception of airborne targets. This timer is of the master-oscillator type, and in complexity is intermediate to the AN/TPS-3 self-synchronous timer just described and the SCR-584 precision timer to be described in Sec. 177. The trigger, gate, intensifier, and deflection waveforms produced by the timer are shown in Fig. 317. They comprise (1) a transmitter trigger, (2) a positive gate for sensitizing the receiver i-f amplifier immediately follow-

ing the transmitted pulse, (3) an initiating trigger waveform for the type B sweep, (4) intensifier pulses for type A, B, and C indicators, (5) type A deflection waveform, and (6) a range marker for the type B indicator, developed in conjunction with the type C intensifier pulse.

The functions of the waveforms mentioned are evident from the previous discussion, with the possible exception of the gate for the type C indicator and the range marker associated with it. We recall from Sec. 17, Chap. I, that the type C indicator

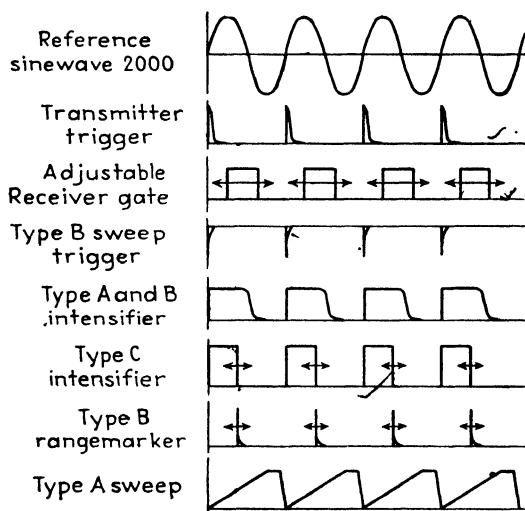



FIG. 317.—Timer waveforms of the SCR-720 radar.

displays range and azimuth information in rectangular coordinates and that the signal-to-noise ratio of this type of indicator is inherently poor because the noise is integrated in the c-r spot during the whole pulse interval. The momentary intensity modulation representing the target must compete against this noise. To minimize the integration of the noise, it is customary to apply a positive intensifier pulse to the grid of the type C indicator, the pulse beginning with the transmitted pulse and lasting only as long as the echo interval of the target under observation. Visual evidence of noise is cut off thereafter by the cessation of the intensifier pulse.

To provide an intensification pulse of the requisite duration, the trailing edge of the intensifier pulse is made adjustable, and from this trailing edge a range marker pulse is developed by

differentiation. The range marker is displayed on the type B indicator in the range coordinate. In adjusting the length of the intensifier pulse, the range marker is moved until it appears on the type B indicator just beyond the target of interest. Thus the intensity of the type C indicator is adjusted to include the target echo but not beyond and thus includes a minimum of background noise.

The type A and type B deflections are initiated simultaneously, and their sweeps are intensified for the same period since both indicate the same extent of the range coordinate. In this equipment the deflection waveform for the type B indicator is developed in the indicator equipment.

The block diagram and simplified schematic of the type SCR-720 timer are shown in Figs. 318 and 319. Consider first the block diagram. The basic timing source is a sine-wave oscillator of the Wien bridge type. The oscillator operates at a frequency equal to the pulse rate, 2,000 pps, and feeds a buffer amplifier. The output of the buffer amplifier is rectified in a half-wave diode and then passed to a two-stage limiting amplifier that develops rectangular waves of 250 μsec duration (one half the period of the 2,000-cps sine wave).

The rectangular waves thus formed are used in two ways. In the first instance, they are passed through a first differentiating stage which produces pulses of shorter duration (15 or 75 μsec) which serve as the intensifier pulse for the type A and type B indicators. In the second case, the rectangular waves control a type A sweep generator consisting of a triode, which charges the sweep capacitor, and a pentode, which assures a constant rate of discharge of the capacitor. The sweep lengths of 15 μsec (maximum range 1.5 miles) and 75 μsec (maximum range 7.5 miles) are selected by switching the sweep capacitor.

The differentiated wave from the first stage is passed through a second differentiating stage that produces a narrow negative pulse to serve as a trigger for the type B deflection circuits. Passage through a third differentiating stage develops a positive pulse of narrow dimension, which is fed through an impedance-matching cathode-follower stage to serve as the trigger for the transmitter.

The transmitter trigger having been established, it is necessary to develop a positive receiver gate pulse, which is delayed for the

duration of the transmitted pulse. This is done by passing the output pulse of the second differentiating stage through an artificial transmission line, which consists of 15 sections, four of which introduce a fixed delay of 1 μsec, the remaining eleven permitting quarter-microsecond increments of delay. The output of this delay line initiates the action of an asymmetrical multivibrator, which develops a positive rectangular wave about 75 μsec long. This passes through a cathode follower to the i-f amplifier grids of the receiver.

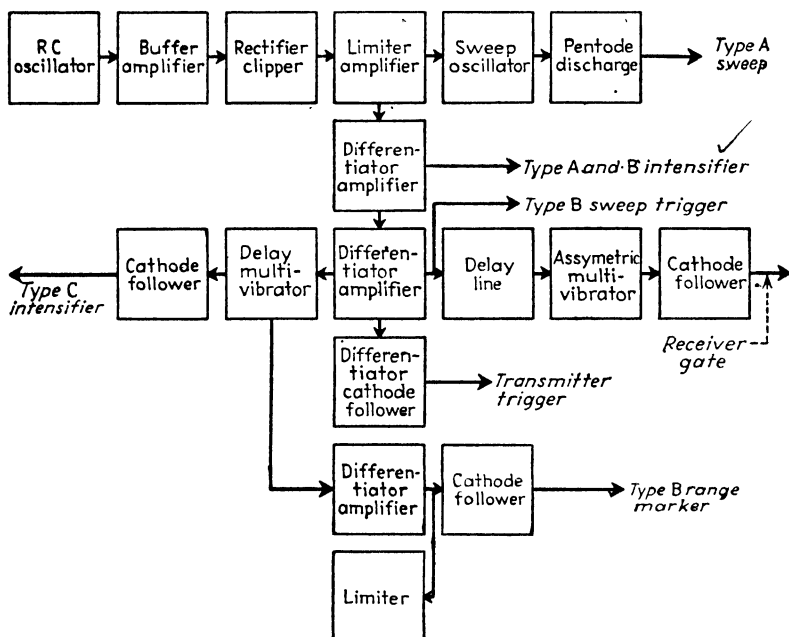



FIG. 318.—Block diagram of SCR-720 timer.

Finally the timer generates the previously mentioned type C intensifier pulse and associated type B range marker. This function begins with the output of the second differentiating stage, which energizes a delay multivibrator. This circuit develops a positive rectangular wave, immediately following the trigger, the length of which is adjustable from 10 to 75 μsec. The rectangular pulse, after passage through a cathode-follower, drives the control grid of the type C indicator and intensifies the sweep for a period which includes the target echo of interest.

The trailing edge of the last-named multivibrator is passed through a differentiating stage to develop the type B range marker. This appears in positive polarity by virtue of the reversing property of the stage. The range marker is applied to the type B indicator grid through a cathode follower.

The schematic diagram, Fig. 319, gives the details of the circuits just described. The Wien bridge oscillator is a double-triode circuit that oscillates by virtue of a feedback path from the plate of the second triode to the grid of the first triode. The feedback occurs through a bridge network that develops a voltage whose phase is most favorable to oscillation at one frequency. The *RC* products in the two upper arms of the bridge are equal to each other and to the inverse of the desired angular frequency of oscillation. The 3-watt filament lamp in one of the lower arms of the bridge stabilizes the amplitude of the output. The lamp resistance increases as the current through it increases, increasing the cathode resistance of the first stage and inhibiting the tendency to increase the amplitude of oscillation. The Wien bridge oscillator is widely used in this and similar applications because of its simplicity, stable amplitude and frequency over a wide frequency range, and purity of waveform. The frequency is readily adjustable by simultaneous adjustment of the resistive elements in the *RC* arms of the bridge.

The buffer amplifier and diode rectifier are conventional. The limiter amplifier employs two tubes that operate at zero grid bias and are overdriven by the high amplitude of the applied waveform. The differentiator stages achieve the desired shortening of the rectangular wave by suitable choice of the series grid capacitor and shunt resistor in each stage. In the first differentiator, a relatively high capacitance (0.0001 μf) in conjunction with a 250,000-ohm resistor produces a differentiated pulse of about 75 μsec duration, for intensifying the longer sweep. Substitution of the 50,000-ohm resistor produces an intensifier pulse about one fifth as long, or 15 μsec, for the short-range sweep. The second differentiator, in contrast, has a 10-$\mu\mu$f series capacitor and develops a trigger of about 1 μsec length.

The delay transmission line is a constant-*k* filter, of 2.5 mh and 25 $\mu\mu$f per section. The variable multivibrator will be recognized from the discussion of Sec. 123. Variation of the

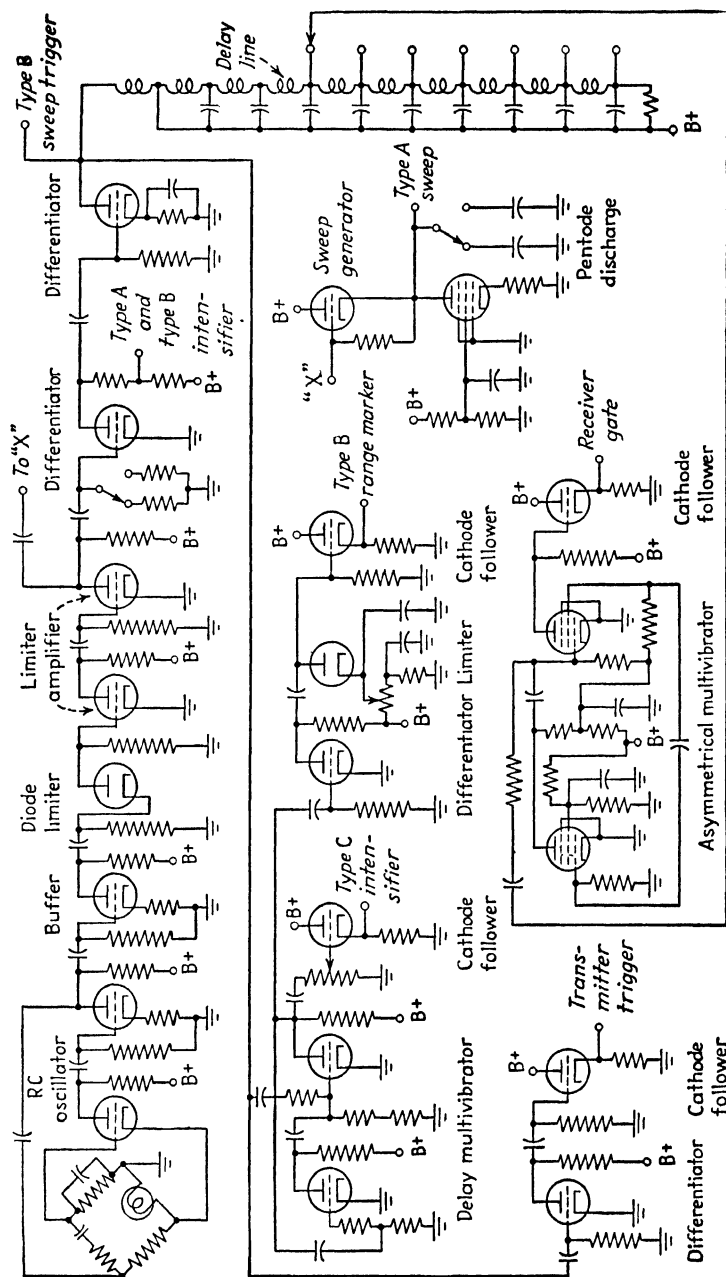



FIG. 319.—Detailed schematic of SCR-720 timer.

type C intensifier pulse is obtained by adjusting the grid resistor of the right-hand triode. The range marker pip is produced by a short-time differentiator (series capacitor 50 $\mu\mu$f) the output of which is limited by a biased diode. All outputs are passed through cathode coupled stages for matching the impedance of the outgoing coaxial lines.

A circuit not previously discussed in detail is the asymmetrical multivibrator that develops the receiver gate pulse. This circuit employs pentode tubes. One pentode acts as a normal amplifier. The other employs the screen grid as the anode for the multivibrator, and delivers the output wave by electron coupling to the succeeding cathode follower. The electron coupling isolates the output loading and permits a high degree of dissymmetry in the multivibrator waveform.

177. Master-oscillator Timers. Crystal-controlled Timing for Precision Range Measurement.—As an example of a timer intended for highly precise measurement, we examine now the synchronization circuits of the SCR-584 radar. This equipment is an outstanding example of the radar art, and will serve as an example of many circuits and structures to be discussed in the following chapters. As a preliminary to the discussion of the timing circuit, therefore, we shall outline briefly its specifications and intended use.

The SCR-584 radar was employed during the war primarily to control the direction and range of 90-mm antiaircraft gunfire. Its specifications, more completely given in Table IX, are as follows: The transmitter radiates 300-kw pulses at approximately 3,000 megacycles (10 cm) from a 6-ft paraboloid that is fitted for conical as well as helical scanning. The pulses are 0.8 μsec long and are repeated at a rate of 1,707 pps.

Two types of indication are provided: type J (circular form of type A) for range measurement, and type ppi for general observation of all targets within range. The maximum range provided on the ppi is 70,000 yd (35 nautical miles), and a shorter range scale of 35,000 yd is also available for observing near-by targets with greater resolution. The type J indicators are two in number, one (the coarse range scope) for measurement of range out to a maximum of 32,000 yd, and a second (the fine range scope) for selecting any 2,000-yd segment of this range. On the screen of this second indicator, the range measurement is

refined by interpolation to a dynamic accuracy of plus or minus 15 yd. The equipment is capable of being reset, on a static target, to a precision of plus or minus 2 yd. In addition to this outstanding range-measurement performance, the radar is arranged to follow automatically in azimuth and elevation any target selected by the operator, the angular accuracy on a moving target being plus or minus 0.03 deg.

The timing equipment of the SCR-584 is called upon to serve the following functions:

1. Establish a pulse repetition rate of 1,707 cps and maintain it so accurately that the echo interval may be read to a timing accuracy of 0.06 μsec, representing a dynamic range accuracy of 15 yd. Crystal control of the repetition rate is essential to meet this requirement.

2. Provide sinusoidal voltages in quadrature at 5,120 cps to deflect the coarse type J scope in its 32,000-yd circular sweep.

3. Provide similar type J sweep voltages at 81.95 kc for the 2,000-yd sweep of the fine range indicator.

4. Provide a transmitter trigger whose time of occurrence can be adjusted relative to the beginning of the coarse type J indicator sweep. This trigger is also used to initiate the sweep of the ppi.

5. Provide an intensifier pulse of 195 μsec duration for the coarse type J range scope.

6. Provide a narrow intensifier pulse (about 3 μsec wide) for the fine range scope. This pulse is also used as a range marker on the ppi indicator to permit selecting a target, from among those displayed on the ppi, for examination and tracking on the fine range scope. This pulse, called the "narrow gate," is also applied to the intermediate frequency amplifier of the receiver to cut off all echo signals except those received during the interval displayed on the fine range scope. This i-f amplifier, one of two in the receiver, controls the automatic following system, as described in more detail in Chap. XI. In a revised version of the SCR-584 equipment, the width of the narrow gate is reduced to less than a microsecond duration, and pulse is dubbed N^2 gate (N^2 = "narrow narrow"). The very narrow pulse permits refining the accuracy of automatic following and mitigates the effects of jamming and confusion among closely spaced targets.

The time sequence of the above-listed pulses and gates is shown in Fig. 320. The time scale in microseconds is shown at the top. The pulses recur at the rate of 1,707 cps, corresponding to a pulse interval of 586 μsec, or an echo interval of 96,000 yd. The echos of interest lie within 32,000 yd (because of the limited range of the 90-mm antiaircraft guns controlled by the radar).

The trigger that controls the transmitter occurs at zero, 586, 1,172 μsec, etc. The transmitted pulse and ppi sweep occur coincidentally. A typical echo, shown at a range of about 13,000

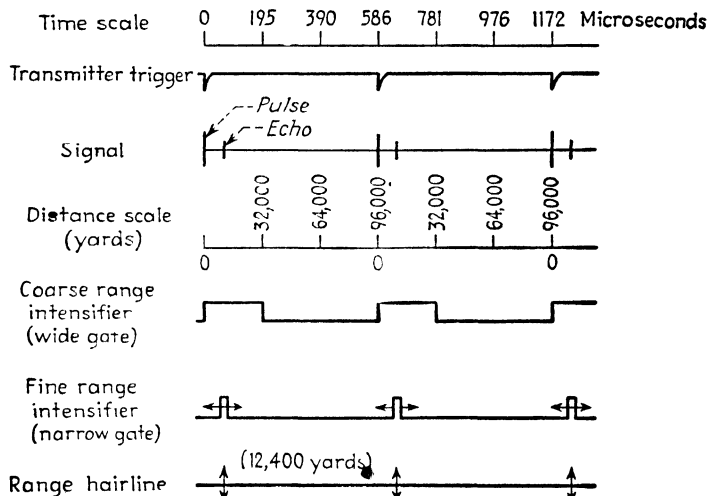



FIG. 320.— Timer waveforms of the SCR-584 radar.

yd, follows each transmitted pulse. This echo is displayed on the coarse range scope, which is intensified by the wide gate pulse of 195 μsec duration (32,000 yd), immediately following each transmitted pulse.

The narrow gate pulse is adjustable in time over the full width of the wide gate. Its position is adjusted by the operator so as to fall under the echo of interest, as shown in the figure. The narrow gate intensifies the fine range indicator sweep during the reception of the echo, controls the gain of the automatic following channel, and serves as an adjustable range marker on the ppi.

The appearance of the type J indicators is shown in Fig. 321. The echo pulses are displayed as radial deflections on the circular sweeps. The coarse sweep, at the left, begins simultaneously

with the transmitted pulse and continues for 195 μsec. The periphery of the tube is marked in the corresponding range units up to 32,000 yd. The fine indicator starts at the beginning of each 2,000-yd interval, but only a portion of one sweep is intensified.

Over the faces of the tubes are hairlines used as pointers. The operator turns a handwheel geared to these hairlines until the hairline on the coarse indicator lies over the target echo of interest, shown in this case at a range of approximately 13,000 yd. The hairlines on the fine indicator, geared to the first through a

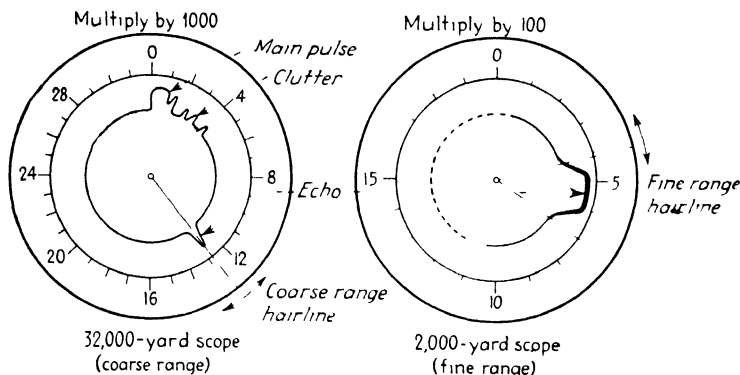



FIG. 321.—Coarse and fine range indicators (type J) of the SCR-584 radar.

16:1 gear train, are positioned at either side of the echo as it appears on the fine indicator. The fine sweep is brightened in the vicinity of the hairlines by the narrow gate pulse. The fine indicator shows that the range exceeds 12,000 yd by 410 yd, that is, the indicated range is 12,410 yd.

It should be remembered that the type J indicators reveal the range of targets only along a particular line of view, that toward which the radiator is pointing. When the radiator is scanning in azimuth, searching for targets, the type J indications change rapidly as the line of view of the radiator passes over the field of view. The type J indicators then give little information of value. When a particular target is selected on the ppi, the radiator is brought to rest when viewing that target, and the adjustable ppi range marker moved until it falls on the target echo. This operation automatically brings the desired echo into view on the coarse and fine range indicators and permits the range to be read at a glance.

Thereafter, the bearing of the target is followed by automatic tracking, and the operator adjusts the range handwheel so as to keep the target echo between the two lines on the fine range indicator. The range is thus followed manually and the range information fed to the associated gunfire computer. The azimuth and elevation information is fed simultaneously from the automatic angle tracking mechanism, described in Chap. XI.

178. Block Diagram of the SCR-584 Timer.—The manner of forming the pulses and gates just described is shown in the block diagram, Fig. 322. The action of the timer starts with the quartz crystal oscillator that operates at a frequency of 81.95 kc, a frequency whose period corresponds to an echo interval of 2,000 yd. The output of the crystal oscillator is fed first to a phase-shifting transformer that develops two sinusoidal voltages of the same amplitude and frequency, but displaced 90 electrical degrees in phase. When these voltages are applied, respectively, to the vertical and horizontal deflection plates of the fine type J indicator tube, a circular Lissajous figure results. Each trace starts at a time determined by the phase of the crystal oscillator and is synchronized thereby with all other timing functions of the equipment. The sweep repeats itself 81,950 times per sec, but is intensified by the narrow gate only for a portion of one sweep in each pulse interval, as previously described.

The crystal oscillator is fed, in the second place, to a trigger generator stage, which selects the peaked wave from the plate of the oscillator tube and sharpens it further. The peaked wave from the trigger generator is then passed to two paths. The first is a series of frequency-dividing multivibrators in cascade. These introduce a frequency division of 48 (4, 4, and 3, respectively), producing at the output a rectangular wave at

$$\frac{81.95}{48} = 1{,}707 \text{ cps.}$$

This is the desired pulse repetition rate. The multivibrator output is not in itself used to trigger the transmitter. Rather, the multivibrator output controls another amplifier, the trigger selector, shown at the lower left.

The trigger selector passes one out of every 48 of the peaked pulses fed to it from the crystal oscillator. Thus the crystal,

FIG. 322. — Block diagram of the SCR-584 timer.

which controls the fine range sweep as previously described, also controls directly the transmitted pulse. The 1,707 cps output of the multivibrator is passed through a succession of shaping and delaying circuits so that it appears at the trigger selector circuit as a positive pulse of 6 μsec duration and of adjustable position. This wave is applied to the screen grid of the trigger selector, allowing it to pass one pulse. The cessation of the 6-μsec pulse thereafter cuts off the trigger selector, blocking the succeeding 47 pulses.

The circular sweep for the coarse range indicator is derived from the output of the second frequency-dividing multivibrator, at 5.12 kc. The rectangular wave output of this multivibrator is passed through a 5-kc amplifier containing a phase-shifting transformer in its output. This transformer develops two sinusoidal voltages of equal amplitude and frequency but separated 90 electrical degrees in phase. These voltages form the circular Lissajous figure, producing three complete sweeps on the coarse indicator screen during the pulse interval. Only the first of these sweeps following the transmitted pulse is visible, being intensified by the wide gate pulse.

The remaining functions of the timer are controlled from the 1,707 cps output of the multivibrator chain. The output is first passed through a clipping stage that steepens the sides of the multivibrator output, to insure exact timing of the succeeding circuits. The clipped wave is then passed, in the first instance, to the narrow gate multivibrator. The circuit generates a positive rectangular wave of adjustable width, the leading edge of which is coincidental with the start of the clipped pulse. Adjustment of the grid resistance of the synchronized grid determines the width of the multivibrator wave, over a range of from 0 to 195 μsec. The trailing edge of this wave initiates the action of the succeeding "gate width" multivibrator that generates the narrow gate itself. Adjustment of the grid resistance in this multivibrator likewise determines the length of the output, that is, the width of the narrow gate. This value is set at approximately 3 μsec. The narrow gate is transmitted to the fine range scope as an intensifier pulse, to the ppi as a range marker, and to the receiver as an i-f gate control.

The second output from the 1,707 cps clipper is passed to the wide gate circuit. The first element of this system is a delay

multivibrator, identical in function to the delay multivibrator in the narrow-gate system. The width of the output is adjustable over the range from 0 to 50 μsec, and the trailing edge of the wave drives the following multivibrator, whose positive output constitutes the wide gate pulse, applied to the coarse range indicator as an intensifier pulse. By adjustment of the wide gate multivibrator, it is possible to shift the intensified portion of the coarse range sweep so that it includes any desired portion of the transmitted pulse.

The third output from the 1,707 cps clipper is the trigger system. The first element is a delay multivibrator similar to the other delay circuits shown. It produces a positive output of adjustable width over 0 to 40 μsec. The trailing edge of this wave initiates the action of the following multivibrator, which delivers a pulse of 6 μsec width, in positive polarity, with leading edge timed to the previous delay multivibrator. This positive pulse is passed through cathode follower and thence to the screen grid of the trigger selector, where it permits conduction through the trigger selector during the 6-μsec period. By adjusting the trigger gate delay multivibrator, the position of the trigger gate pulse may be moved relative to the triggers fed to it from the crystal circuit. It is thus possible to select any one of three successive triggers. The succeeding 47 triggers are blocked. The trigger selector also sharpens the triggers as they pass through it. The end result is a series of accurately timed triggers at 1,707 cps, with sharp leading edges rigidly synchronized with the crystal control and the range indicator sweeps. These triggers initiate the transmitted pulses as well as the ppi sweeps.

179. Simplified Schematic of the SCR-584 Timer.—Figure 323 shows a simplified schematic diagram of the timer just described. The component elements will be recognized, for the most part, by comparison with the circuits analyzed in Chap. VI. The crystal oscillator is a conventional pentode oscillator, whose plate circuit contains the primary of the phase shifting transformer. A second winding of this transformer is tuned and loaded with resistance that can be adjusted to balance the amplitudes of the two phases. A third winding, coupled to both the others, provides the quadrature voltages against ground.

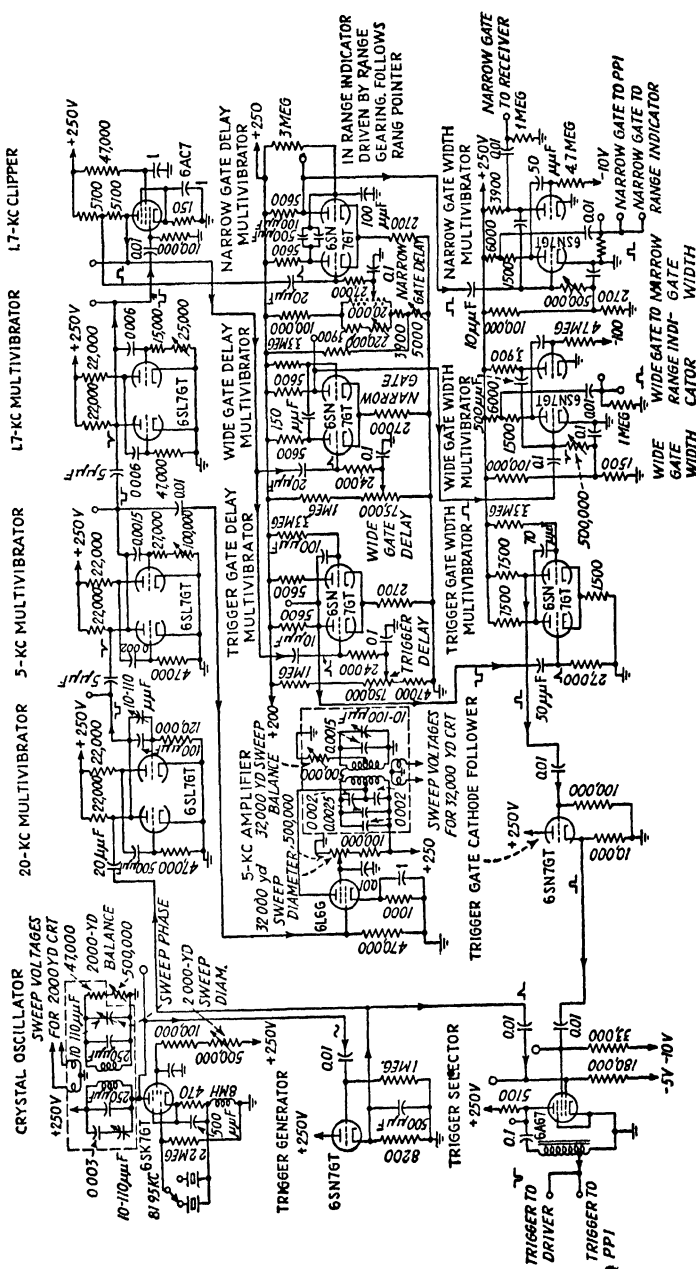



FIG. 323.—Schematic of the SCR-584 timer.

A connection at the plate of the oscillator leads directly to the cathode-coupled trigger generator. This circuit sharpens the pulses by virtue of the short time constant (4.1 μsec) of the *RC* shunt elements in series with the cathode. The period of the 81.95 kc wave is 12.2 μsec. The cathode passes increasing current during the first quarter cycle of the applied wave, for about 3 μsec. When the applied voltage starts downward during the second quarter cycle, the cathode voltage is not changed at once, due to the charge on the cathode capacitor. Hence the tube is cut off before the end of the first cycle and remains cut off for the following half cycle. The voltage appearing across the cathode capacitor, which is fed to the succeeding circuits, is peaked by this succession of events, as shown in the figure.

The chain of three multivibrators that divides the oscillator frequency will be recognized as slightly asymmetrical versions of the basic circuit described in Sec. 123, Chap. VI. The trigger synchronization is applied to the grid of the right-hand triode. The multivibrator constants are adjusted to produce a free-running frequency slightly lower than the desired value of 20.49 kc. Every fourth synchronization pulse drives the right-hand triode grid to conduction, but intervening pulses have no effect due to the depressed condition of the opposite grid. Rectangular waves appear across the plates of the two tubes, that from the left-hand tube being passed to the succeeding divider. The time constant of the right-hand grid in each case is adjusted by a variable capacitor or resistor.

The clipping stage operates by virtue of grid-current biasing, which flattens the negative peaks of the plate waveform and similarly flattens the positive peaks of grid voltage cutoff. The gate delay multivibrators that follow are alike in action. Each consists of a pair of triodes coupled by a common cathode resistance, as well as by a connection from the left-hand plate to the right-hand grid. No coupling connection is provided between plate and grid in the opposite direction, however, since the cathode coupling completes the feedback path.

The wave from the clipper is passed first through a differentiating circuit (*C* 10–20 $\mu\mu$f, *R* 24,000 ohms) in the left-hand grid of each multivibrator. The positive differentiated wave resulting from the leading edge drives the left-hand triode to

conduction, and this condition persists for a time determined by the time constant of the *RC* feedback from the plate of that tube to the grid of the other. The grids of both tubes are returned to the *B* plus to assist in quick discharge, and the time of recovery is determined by adjusting the amount of this positive bias on the left-hand grid in each case. An asymmetrical waveform results.

The trailing edge of the positive half of this wave establishes the timing reference of the succeeding gate-width multivibrators These are conventional balanced circuits of the asymmetrical type. In the narrow-gate and wide-gate circuits, the width is adjusted by the grid return resistors in the left-hand triode in each case. The trigger gate-width circuit is fixed.

The succeeding cathode-coupled stage is conventional; its purpose is to avoid capacitive loading on the preceding plate, as well as to feed the succeeding screen grid from a low impedance. The trigger selector is a pentode, with triggers applied to its control grid and trigger gate to its screen grid. Only when both pulses are present simultaneously will the tube conduct and apply an output trigger to the plate circuit. The plate contains an autotransformer (pulse transformer type) that reduces the output impedance to a low value, to match the succeeding coaxial cables to the transmitter and ppi circuits.

The remaining section of the timer schematic is the 5-kc amplifier that develops the sweep for the coarse range indicator tube. This stage employs a beam power tube, the voltage on whose screen is adjusted to vary the amplitude of the output voltages and thus to control the diameter of the circular sweep. The plate output is passed to a tuned transformer, balanced with respect to ground. The two voltages in quadrature appear across a third winding, as in the case of the fine range sweep circuit.

180. Other Timing Methods.—The timing methods thus far described (self-synchronous spark gap, sine-wave oscillator at pulse frequency, and crystal-controlled oscillator plus frequency division) are employed in the majority of radars. There are, however, many variations of these systems. A method used in early designs (for example, SCR-268), and still considered useful, is that illustrated in the block diagram of Fig. 324. This timer employs an *LC* oscillator at pulse frequency (4.098

cps), and a continuous inductive phase shifter for range measurement.

The oscillator output is conveyed through two chains. One develops the transmitter trigger, the other the range sweep. The first chain starts at the plate of the oscillator, where a slightly flattened wave appears. This is passed to a limiter amplifier that develops steep edges. These edges, passed through a differentiator network, appear as two narrow pulses of opposite polarity. The first of these, positive in direction, is passed through a biased amplifier that removes the succeeding

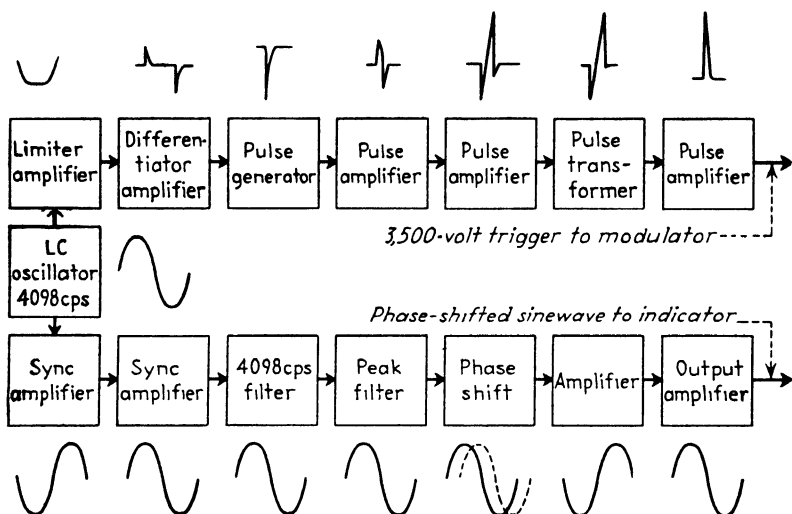



FIG. 324.—Block diagram and waveforms of the SCR-268 master oscillator timer.

(negative) pulse. Further amplification of the first pulse and passage through a pulse transformer result in a pulse of about 5-μsec width, based on a negative bias of $-3{,}000$ volts and extending to $+500$ volts. This is the keying impulse which drives the modulator.

The second chain from the timing oscillator is taken from the tank circuit in sinusoidal form. Two stages of linear amplification are followed by a narrow bandpass filter (4,098 pps + 20 per cent) designed to remove harmonics. The purified sine wave is then split into two phases, 90 electrical deg apart, and fed to the two stator coils of an inductive phase shifter (Helmholtz coil). As previously described, in Sec. 141, this arrangement

produces a rotating magnetic field that induces a sine-wave voltage in the rotor coil. The position of the rotor coil determines the phase of the induced voltage, relative to that applied to the stator coils through the phase splitter. The phase may be continually advanced or retarded by rotation of the rotor element.

The phase shifted wave is passed to the sweep generator circuits, which develop sawtooth waves of total period 244 μsec (the period of the 4,098 cps wave). The time of occurrence of these sawtooth waves can be adjusted relative to the transmitted pulse and tle echoes, and this adjustment causes the pulses to move on the type A indicator screen. The echo pulse of interest is shifted under a reference hairline, and the range to that echo read from the calibrated dial of the phase shifter.

The manner of fulfilling the basic functions of the timer, enumerated at the beginning of this chapter, is fundamentally dependent on the specifications of the radar, its intended use, and the accuracy demanded of its indications. Even when these quantities are fixed, considerable latitude remains in the choice of specific circuits. The foregoing examples must not be considered as inclusive, but merely as suggestive of the many synchronization schemes available.

CHAPTER IX

TRANSMITTERS AND RADIATORS

The second group of elements we shall consider comprises the transmitter, transmission-line system, and radiator. The transmitter is generally divided into two parts: the keyer or modulator, which establishes the duration and interval of the pulses; and the r-f (radio-frequency) oscillator, which transforms the modulating pulse into a corresponding train of r-f oscillations. The output of the transmitter is conducted by an appropriate transmission line or waveguide to the radiator.

Each of these elements is composed of component circuits and structures, which have been considered separately in Chaps. VI and VII. The particular form of the combination depends so largely on the intended use and physical dimensions of the radar that the general principles of design are best illustrated by specific examples. There are, however, certain trends, particularly in modulator design, that can be discussed in a general way before embarking on specific descriptions of typical transmitter equipment. Modulator circuits have received concerted attention because they expend a large fraction of the primary power required by the radar.

181. Functions and Requirements of the Modulator.—The modulator is required to fulfill the following general functions:

1. To accept the basic trigger pulse provided by the timer and utilize it without appreciable delay or variation.
2. To produce a modulating pulse of the requisite duration and thereby to control the r-f pulse width.
3. To produce a pulse of proper shape (usually rectangular), with leading and trailing edges meeting specified times of rise and fall.
4. To supply the modulating pulse at the proper voltage to secure optimum efficiency and stability in the modulated r-f oscillator.
5. To match the impedance of the r-f oscillator during the pulse. This implies that the current flow, at the required voltage

specified above, shall be that demanded by the r-f oscillator at the optimum operating point.

6. To meet all the foregoing requirements with maximum economy in the use of primary power, space, and weight.

To achieve these ends, four basic circuit elements are required: a power supply, a circuit element or elements for energy storage between pulses, an electronic switch for discharging the stored energy through the r-f oscillator during the pulse, and a control system for turning the switch on and off at the requisite rate and with the necessary suddenness (to produce steep leading and trailing edges in the r-f pulse).

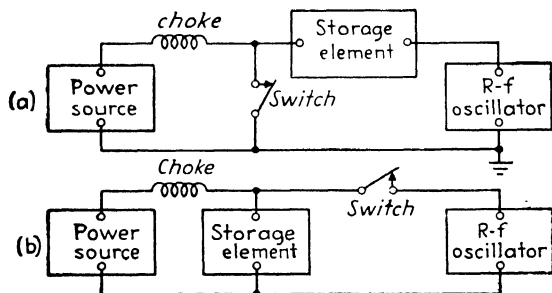



FIG. 325.—Basic modulator circuits: series charging (*a*) and shunt charging (*b*).

These elements are connected in one of the prototype circuits shown in Fig. 325. In general, two circuit loops are required, with a common arm. This common arm may contain the switch, or it may contain the storage element. In the first case (series charging, Fig. 325*a*) the storage element is charged in series with the r-f oscillator and is discharged through the oscillator by closing the switch. In the second case (shunt charging, Fig. 325*b*) the storage element is shunted across the power source and the switch completes the circuit through the oscillator load. In both cases, an impedance (series resistance or inductance) must be connected in series with the power supply to prevent its being short-circuited during the pulse. The two schemes are widely used, and the choice between them rests on individual preference. One advantage of the series charging system is the fact that one terminal of the switch and the r-f load can be grounded, a convenience when a rotary spark gap is used.

182. Types of Modulator Circuit Elements.—The power source of a modulator must be of the high voltage variety *because the r-f oscillators used in radar are inherently high voltage devices.* The triode oscillators require high voltage to secure short transit time and full emission without excessively small spacing between elements. The cavity magnetron is most efficient when designed for high voltage operation.

The use of high voltage has serious disadvantages, particularly in airborne equipment, because of the limitations imposed by corona and insulation breakdown. Some reduction of the voltage level in the modulator proper may be obtained by the use of a

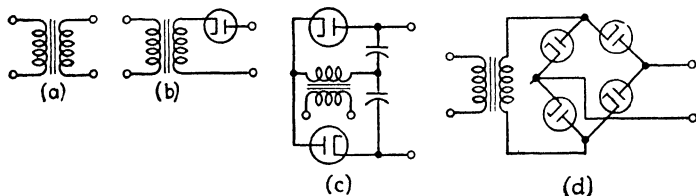



FIG. 326.—Modulator high-voltage power supplies: (*a*) transformer for synchronous circuit, (*b*) half-wave rectifier, (*c*) full-wave rectifier, and (*d*) bridge rectifier.

step-up pulse transformer between modulator and oscillator, but the step-up ratio is usually limited to not more than 4:1.

The power source may be of the a-c type if the primary power frequency is equal to, or an integral multiple of, the pulse rate. This system is not suitable when the primary power frequency is 60 cps, since such a low pulse rate is generally not suitable. The majority of modulator power supplies employ rectification in some form or other, that is, the storage element is charged from a unidirectional source. Voltage doubling arrangements are often used to obtain higher voltage from a given supply. Typical modulator power sources are shown in Fig. 326.

The storage element of the modulator may be a capacitor, an inductor, or an artificial transmission line composed of capacitive and inductive elements. The simple capacitor is used only when the pulse is terminated, as well as initiated, by the switch. The voltage across the capacitor tends to drop as the charge is withdrawn from it during the pulse, and this decrease appears as a slope in the top of the pulse. To secure a flat-topped pulse a large capacitor must be used and only a small fraction of the

charge removed during each pulse, that is, the pulse must be cut off, by the switch, while the capacitor remains substantially fully charged.

In cases where the switch cannot terminate the pulse, the storage element is usually an artificial transmission line. The simple constant-k low-pass filter is commonly used. The inductive elements tend to support the current drop as the capacitors are discharged and the result is a pulse whose top remains nearly flat up to the point at which the capacitors are completely discharged. When the pulse must have a top of maximum flatness and the size and weight requirements are severe, an unorthodox filter section designed by Guillemin (Guillemin line) is often used.

The switching element of the modulator is perhaps the most critical element. It must conduct currents of tens or hundreds of amperes at voltages in the tens of kilovolts. It must close the modulating circuit with great speed (not over 0.1 μsec is allowable in the time of rise of cavity-magnetron modulating pulses). In some cases the switch must terminate the connection with nearly equal suddenness. It must, in the case of master-oscillator timing systems, perform the switching in rigid synchronism with a considerably less powerful control pulse. In self-synchronous systems the variation in pulse timing must not be excessive (for example, not over 100 μsec). Finally, the switch must not consume an appreciable part of the power passed through it. Between pulses, it should preferably consume no power at all.

These requirements have been met by three types of switch devices: the spark gap[1] (fixed and rotary); the high vacuum electron tube ("hard" triodes and tetrodes); and the gas-filled tube (thyratron). All three devices have been used widely, the hard tube in the early designs, the rotary spark gap in self-synchronous equipment, and the hydrogen-filled thyratron.

The timing function of the rotary spark modulator has already been discussed in Chap. VIII, Sec. 174. Our concern here is its function as a switch, that is, its current-carrying and insulating properties and its useful operating life. A typical rotary gap is shown in Fig. 327. The arc that passes between rotor and

[1] GOUCHER, F. S., and others: Spark Gap Switches for Radar, *Bell Sys. Tech. J.*, **25**, 563 (October, 1946).

stator electrodes can carry any current required, up to several thousand amperes, but the heavy current tends to eat away the electrodes if they are made of low-melting-point material. Tungsten, a high-melting-point metal, is virtually the only material used for this purpose. An indirect corrosive effect is produced by the formation of ozone by the spark and by transformation of water vapor into nitric acid. Enclosing the gap in a pressure-tight container is advisable to assure reliable operation. The shielding provided by the housing is desirable to prevent direct pickup of r-f energy generated by the spark. Properly designed rotary gaps will operate upwards of 1,000 hr without the necessity of changing the electrodes.

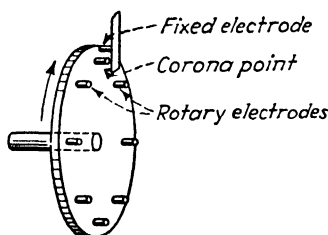



FIG. 327.—Rotary-spark-gap modulator switch.

The start of the modulating pulse in a rotary gap modulator is produced by the breakdown of the air as the rotor and stator electrodes approach one another. The termination of the pulse, however, is not produced by the fact that the electrodes subsequently draw apart. The arc, once started, would continue until the electrodes were separated an inch or more, if the voltage across them were not reduced to a low point at the termination of the pulse. The voltage reduction is provided by the discharge of the artificial transmission line. The line becomes completely discharged after the voltage wave completes two traverses of its length. The arc is thereby cut off and the electrodes move apart while the pulse-forming line is recharging from the power source.

The necessity of sudden voltage reduction at the end of the pulse makes it impractical to employ a single capacitor as the storage element in the rotary gap modulator. While a single capacitor, of suitably small capacitance, might be completely discharged during the pulse length, the shape of the pulse would be exponential rather than rectangular. A larger capacitor, capable of producing a flat top during the pulse, would tend to prolong the pulse, since the arc cannot extinguish itself.

Fixed (nonrotary) spark gaps have been the subject of much experimentation, but open gaps are not used because the timing of the pulse (determined by an auxiliary trigger electrode)

is critically dependent on the separation between the current-carrying members. These members are subject to severe corrosion that gradually increases the separation. The operating life is a few hundred hours at most, which is not considered sufficient for practical application. The hydrogen thyratron, and fixed gaps within a low-pressure atmosphere of hydrogen and argon are satisfactory substitutes wherever the timing variations of the rotary gap cannot be tolerated.

The second type of switch, the high vacuum tube, differs from the spark gap and thyratron in one important particular: its control electrode (control grid) can cut off the modulating current when the driving impulse is removed. This means that the width of the driver pulse must be equal to the duration of the desired modulating pulse. A driver stage, which develops a control pulse of the proper width, is generally required when a hard tube is used as the modulator switch. By the same token, the storage element does not determine the pulse duration, and hence it can be of the single-capacitor type.

The principal disadvantage of the hard tube is its low efficiency. The high current passed through the modulator must be provided by thermionic emission and this requires, even with highly emissive cathodes, a large and steady supply of cathode-heating power. Moreover, the voltage drop through the tube during conduction is appreciable (10 to 20 per cent of the power supply voltage) and this drop is subtracted from the useful voltage output of the modulator. Despite these power losses, the hard tube is much used in precision radar equipment, where exact timing of the pulse is required. The tube types most used are triodes with heavy thoriated filaments. The construction is generally not different from conventional triode transmitting tubes. Oxide-coated cathodes are employed in tetrode modulators. These tubes are constructed to permit high voltage operation and to avoid electron emission from the control grid. Table XII lists typical hard-tube modulators.

The remaining modulator switch is the thyratron, which possesses the heavy-current capacity of the rotary spark gap and permits accurate timing without rotary motion. When thyratrons were first applied to modulator service, the only types available had serious drawbacks. The mercury thyratron was capable of operation at high voltages but required tempera-

ture control to assure reliable operation, since the mercury vapor pressure changes over wide limits within the temperature range encountered in service. This proved a serious difficulty in airborne equipment. The other available thyratrons, containing

TABLE XII.—HIGH VACUUM TUBES USED IN MODULATOR SERVICE
Triodes (Thoriated Tungsten Cathodes)

Type no.	Cathode heating, volts, amp	Peak emission, amp	Maximum anode voltage, volts	Average anode dissipation, watts	Maximum duty cycle
304TL	5, 26 10, 13	1	3,000	300	
304TH	5, 26 10, 13	0.9	3,000	300	
6C21	7.5, 16	15	30,000	300	0.002
1,000UHF	7.5, 16	0.75	6,000	1,000	
Tetrode (Oxide-coated Cathode)					
5D21	26, 2.1	15	20,000	60	0.001

inert gas (argon or helium) at low pressure, were limited to low-voltage applications, since they cannot withstand an inverse voltage of more than a few thousand volts. These tubes were used in some low-power designs. The use of thyratrons did not become widespread until the development of the hydrogen-filled type. Table XIII gives the characteristics of several thyratrons used in modulator service.

TABLE XIII.—THYRATRONS USED IN MODULATOR SERVICE

Type no.	Cathode heating, volts, amp	Peak anode current, amp	Peak forward voltage, kv	Max pulse rate per sec	Average anode current, amp
3C45	6.3, 2.1	35	3.0	2000	0.045
4C35	6.3, 6.0	90	8.0	4000	0.100
5C22	6.3, 10.6	325	16.0	....	0.200

The hard tube and the thyratron tube, being controlled by master-oscillator timing systems, require some sort of driving pulse. In the hard-tube circuits, the driver pulse must have steep leading and trailing edges. The thyratron requires only a sharp leading edge and depends on a sudden drop in anode voltage (produced by a pulse-forming line) to terminate the pulse. Several pulse-driving circuits are described in Sec. 184.

183. Basic Modulator Circuits.—The prototype circuits shown in Fig. 325 take several well-defined forms in practice, depending on the type of power source, switch, and storage element employed. The circuits divide conveniently into two groups,

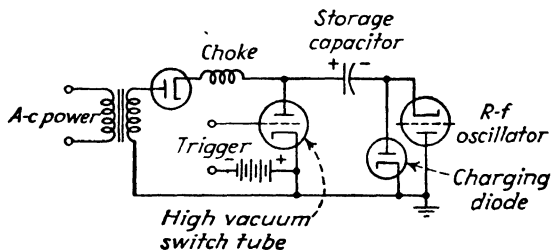



FIG. 328.—Hard-tube modulator circuit using single storage capacitor.

those employing a single capacitor with a hard-tube switch, and those employing a pulse-forming line and a spark gap or thyratron switch. The first type is shown in Fig. 328. This circuit will be identified as the series-charging type that is most widely used in hard-tube modulators. The operation of the circuit is as follows: The storage capacitor is charged, between pulses, from the d-c power source through the series circuit formed by the charging diode, the inductance, and the power source. The hard-tube switch and r-f oscillator are inactive during this interval and represent open circuits.

When the driver applies the positive driving pulse to the hard-tube switch, the grid bias is overcome and the switch tube suddenly conducts, reaching full conduction in a fraction of a microsecond. The inductance L prohibits any sudden increase of current from the power supply, and thus the anode current of the switch tube is supplied entirely by the capacitor. The direction of the current during the discharge phase is the reverse of the charging direction, so the charging diode does not conduct. Consequently, the discharge current from the capacitor is passed

entirely through the oscillator, and the r-f pulse is thereby generated.

At the conclusion of the driver pulse, the hard-tube switch suddenly becomes nonconducting and the discharge ceases. The capacitor has lost but a small fraction of its total charge and this amount is replaced, through the charging path, during the succeeding interval between pulses.

This elementary explanation omits an important circuit element that assumes an active role during the pulse. This is the total shunt capacitance between the switch tube anode and ground, consisting of stray and wiring capacitance and the capacitance of the r-f load. This shunt capacitance may amount to 50 $\mu\mu$f or more in typical circuits and this value is high enough to limit the steepness of the trailing edge of the pulse. The time of rise of the leading edge is determined by the shunt capacitance and the internal plate resistance of the switch tube, which is generally low enough to permit the pulse to rise within 0.1 μsec. The time constant for the trailing edge, on the other hand, involves the resistance of the r-f load, which is generally large enough to prolong the trailing edge a microsecond or more, far too long if the pulse width itself is only 1 μsec This prolongation may seriously affect the earliest time at which the receiver can see echoes, since even a very low amplitude portion of the transmitter trailing edge can be seen after the t-r box has ceased conduction.

To provide a steep trailing edge it is customary to insert a small "peaking" inductance (Fig. 329), so chosen that it resonates with the total shunt capacitance at a frequency equal to the inverse of twice the pulse width The oscillatory current set up in this resonant circuit accelerates the discharge of the shunt capacitance, and thereby steepens the trailing edge of the pulse. A time of fall of 0.4 μsec or less can be achieved by this method.

One consequence of inserting the inductance is the fact that the oscillations set up in the LC circuit may be prolonged over several cycles. If this occurs, the positive halves of the cycles may excite the r-f load at intervals after the main pulse. These residual oscillations are damped out by a diode which absorbs energy during the negative excursions. This diode is connected in the same position and polarity as the charging

diode of Fig. 328, but it no longer serves as a charging element, since the inductance provides a charging path of lower impedance during the pulse interval.

The circuit shown in Fig. 329 is often combined with a pulse transformer and transmission line, which permit separating the r-f components from the modulator components. The transformer winding between the r-f oscillator cathode and ground is commonly wound in two separate parallel sections, as shown in Fig. 330. These windings provide separate paths for the cathode-heating current, and avoid the necessity of providing a cathode-heating transformer with high-voltage insulation between windings. The total shunt capacitance to ground is also reduced thereby.

FIG. 329.—Insertion of peaking inductance to sharpen trailing edge of modulator pulse.

The second type of circuit, employing a pulse-forming line, displays several variations in respect to the power source. Alternating-current or d-c sources may be used, and each may employ resonance effects to increase the peak voltage attained

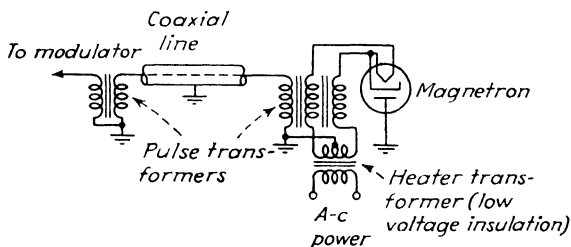



FIG. 330.—Pulse transformers and coaxial line to permit separation of modulator and r-f oscillator.

just prior to discharging the storage element. These circuits are shown in Figs. 331 and 332. In Fig. 331*a* the simplest type is shown, the d-c resistance charging circuit. The pulse forming line is charged from the d-c power source through a resistance. The maximum voltage to which the line can be charged is equal to the power source voltage. The *RC* time constant formed by the capacitance of the pulse-forming line and the charging

resistance must be long compared with the pulse width, but short *compared with the pulse interval.*

The principal disadvantage of this charging method is its low efficiency, since at least one half of the energy derived from the power source is dissipated in the charging resistor. Since the energy storage in the line is equal to or greater than the energy in the transmitted pulse, the d-c resistance charging method cannot offer an efficiency higher than 50 per cent, and it is therefore seldom used.

When resonance is added to the circuit by the insertion of a charging inductance of the proper size, higher efficiency may be

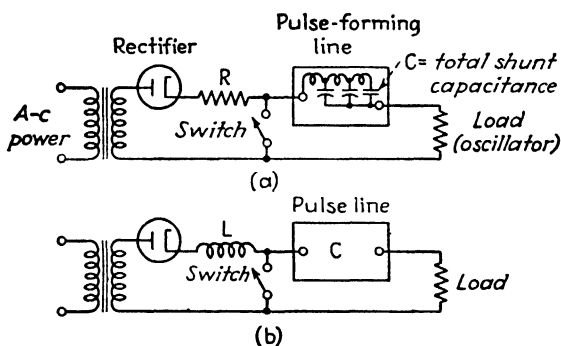



FIG. 331.—D-c charging modulator circuits: (a) series-resistance choke, and (b) inductive choke.

achieved. The circuit (Fig. 331*b*) is widely used in practice. The inductance is so chosen that its period of resonance, with the line capacitance, is twice the pulse interval. When the switch (spark gap or thyratron) is first closed, the circuit is thrown into a damped oscillation. The first peak of this oscillation occurs at the conclusion of the succeeding pulse interval. Moreover, the voltage across the pulse line at this peak is nearly twice as great as the power source voltage. Correspondingly lower voltage components may be used, therefore, in the power source.

Each opening of the switch excites a similar half-wave sinusoid, which is abruptly terminated by the successive closing of the switch. The efficiency of this circuit is 100 per cent if the inductance and power supply have negligible resistance. In practice the resistances may be kept small enough to achieve 95 per cent efficiency.

In radar systems that must operate with adjustable pulse widths or pulse intervals, correspondingly adjustable values of the pulse-line capacitance and inductance must be provided to meet the pulse-forming and resonance-charging requirements. If the space and weight of the extra components cannot be tolerated, a diode may be inserted in series with the charging inductance. This prohibits reverse conduction and allows the capacitor to maintain a peak voltage of nearly twice the power source voltage. Thus the necessary high voltage is available over the pulse interval even if the particular values of inductance

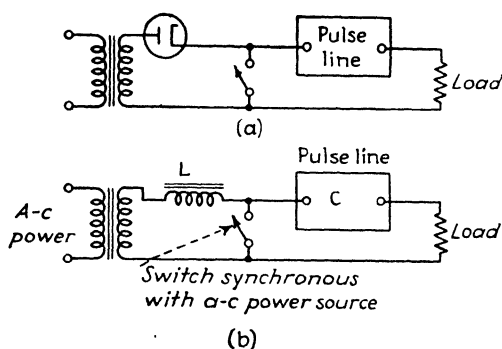



FIG. 332.—Synchronous modulator circuits: (a) with series diode, and (b) with LC resonant circuit.

and capacitance deviate somewhat from the values prescribed for best operation.

An a-c power source may be used with a diode for resistance charging (Fig. 332*a*). This is essentially the equivalent of a d-c supply, but the efficiency is increased since the losses in the diode are smaller than those of the charging resistance. Better than 75 per cent efficiency is usually achieved. However, there are stringent limitations on the primary power frequency. The switch must close while the diode cannot conduct, that is, during the negative halves of the power cycles. This means that the pulse rate and the power frequency must be integrally related, and the power frequency must be equal to or some exact whole multiple of the pulse rate. Since pulse rates lower than 100 per second are seldom suitable, the power frequency must be in the hundreds, or thousands, of cycles per seconds. This type of power source is convenient only for light-weight (for example, airborne) equipments.

Resonance charging may also be used with an a-c power source (Fig. 332*b*). The inductance and pulse-line capacitance are chosen to resonate at the power frequency, so a current of steadily increasing amplitude tends to flow through the circuit. If the series resistance of inductance and power supply are low, the second half cycle of current may produce a voltage across the capacitance that exceeds the peak-power source voltage by a factor of 3. If the discharge switch is closed at the peak of this second half cycle, a correspondingly high voltage pulse may be developed. The same limitations on power frequency

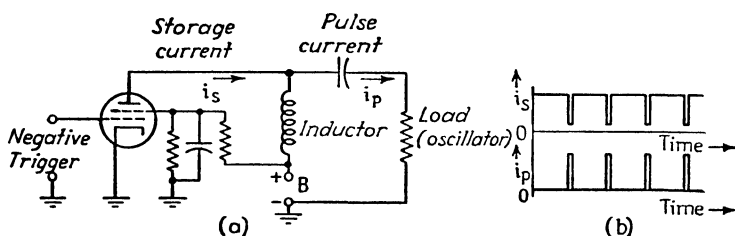



FIG. 333.—Use of inductance as a storage element.

apply as in the case of the diode a-c charging circuit. The more power cycles per pulse interval, the higher the voltage attained.

When the power source voltage is limited to low values (1,000 volts, approximately) the foregoing methods of charging the storage element are not satisfactory. In such cases, inductive storage[1] may be used. High peak voltage is achieved by suddenly interrupting the current through an inductance just before the transmitted pulse, and by closing the modulator switch immediately thereafter, while the impulse voltage still persists across the pulse-forming line. A typical circuit and its waveforms are shown in Fig. 333.

184. Driver Circuits.—Hard-tube and thyratron modulators require synchronizing circuits for initiating and (in the hard-tube case) terminating the pulse. Since the bias that withholds conduction through the switch tube between pulses is generally several hundred or a thousand volts, the synchronizing pulse must be of correspondingly high magnitude. The circuits providing the control pulse with the necessary amplitude and steepness are known as "driver circuits."

[1] PETERSON. E., Coil Pulsers for Radar, *Bell Sys. Tech. J*, **25**, 603 (October, 1946).

The driver circuit is essentially a power amplifier that accepts the trigger pulse from the master-oscillator timer, shapes, and amplifies it. The driver generally consists of two or more stages and the problem of operating these stages at high efficiency is an important one. In the interest of efficiency it is highly desirable that all high-current stages pass current only during the pulse, which implies no phase reversal between cascaded stages. Pulse transformers may be used to change polarity, or cathode coupling may be used.

Driver stages, like timers, may be designed variously to meet the particular requirements of the modulator. Three types of

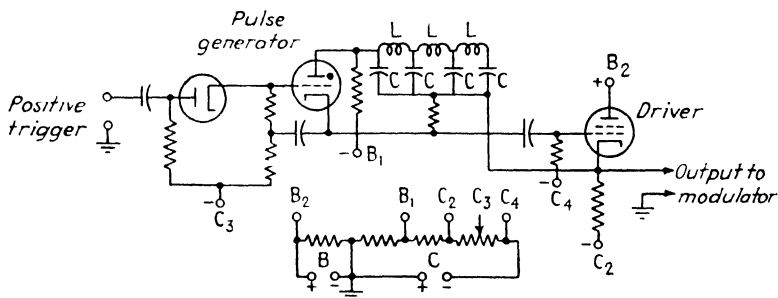



FIG. 334.—Bootstrap driver circuit.

driver that have achieved wide use will be described: the bootstrap driver, the line-controlled blocking-oscillator driver, and the saturable reactor driver.

The bootstrap driver circuit is shown in Fig. 334. The input accepts a trigger pulse from the timer and passes it to a low-level thyratron pulse-forming circuit. The output of the thyratron stage is passed to the grid of the power amplifier stage, an oxide-cathode double tetrode. This stage is, in turn, cathode coupled to the grid of the modulator stage. This driver produces a steep-sided rectangular pulse of controlled duration which is well suited to drive hard-tube modulators (which must be stopped as well as started).

The simplified schematic (Fig. 334) shows a diode in series with the grid of the thyratron stage. If this diode were not present, the grid would be driven below the bias value by the negative portions of the applied waveform, and this negative potential would attract positive ions. Bombardment of the

grid by these ions may exceed the grid dissipation limit and shorten the life of the thyratron.

The thyratron stage itself is very similar to a thyratron modulator operated at low power level. The input end of pulse-forming line is connected between the anode and the cathode, and is open circuited at the opposite end. Between pulses this line is charged through the high resistance (0.6 megohm). When the trigger pulse is applied to the thyratron grid, conduction occurs and the pulse line is effectively short-circuited. A voltage wave, thereby set up, travels down the line and is reflected at the open circuit without change of sign. On returning to the anode end of the line, the wave has removed the charge from the line, the thyratron anode voltage drops to zero, and conduction ceases. The positive trigger on the grid has by this time disappeared so the grid regains control and the tube remains nonconducting until the next trigger is applied. The L and C constants of the pulse-forming line are chosen, in accordance with Eq. (430), to produce the desired pulse width. Sections of the line may be inserted or removed by a switch to alter the pulse width.

The output pulse of the thyratron stage is developed across a cathode resistor and passed to the grids of two beam tetrodes in parallel. The pulse drives the tetrodes into the region of grid current, and thus the top of the pulse is flattened by grid-current limiting.

The grid of the tetrode stage is connected in bootstrap fashion, that is, to the cathode, so the tetrode grid and all connecting circuit elements are raised in potential during the pulse. To insure steep edges, the capacitance of these elements to ground must be minimized, and this is accomplished by enclosing the elements in a shielded container and by employing a cathode-heater transformer with low interwinding capacitance.

The power supply of the bootstrap driver is arranged to provide the bias for the modulator, that is, the 1,000-volt output pulse is superimposed on a bias of -700 volts relative to ground.

The line-controlled blocking-oscillator driver is shown in Fig. 335. It consists of a single tube (a double tetrode, type 829A) connected as a blocking oscillator with a pulse-forming line in shunt from the grid winding of the feedback transformer to ground. Between pulses the line is charged by the grid bias

supply of the driver stage. A positive trigger, applied to the grid end of the line, is passed through the first capacitance of the line (before reflection occurs) directly to the grid of the blocking oscillator. The positive feedback through the transformer between grid and plate causes the grid to be driven rapidly positive and the tube conducts a heavy current that tends to reach an equilibrium value. Grid-current blocking is not at once effective, that is, the oscillator does not cut itself off in the manner of the conventional blocking-oscillator circuit. Instead, voltage induced in the grid winding by the sudden conduction of anode current is applied to the line, travels to the open-circuited end, and is reflected. Upon completion of the reflection the line becomes charged positively by an amount

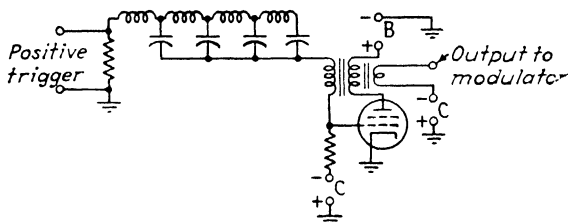



FIG. 335.—Line-controlled blocking-oscillator driver.

that overcomes the positive voltage on the grid, reducing the grid somewhat below cathode potential. This fall of grid voltage is fed back regeneratively and the grid is thereby driven sharply to cut off, terminating the pulse. A tertiary winding on the transformer delivers the resulting pulse to the succeeding modulator stage, in positive polarity, superimposed on an appropriate negative bias.

It is evident that the blocking oscillator driver is a considerably simpler circuit than the bootstrap driver previously described. The blocking oscillator circuit is, in fact, a more recent development made possible by the three-winding pulse transformer, a component not available when the bootstrap circuit was designed.

The saturable reactor circuit[1] is shown in Fig. 336. It is particularly adapted to low-voltage sources. The circuit charges the storage capacitor between pulses, while driver tube and oscillator are inactive. The saturable core reactor, when driven past

[1] See reference, page 470.

its saturation limits, acts as a switch and closes the circuit to the oscillator. Part of the inductive element of the circuit is composed of a saturable core reactor (Sec. 129, Chap. VI). The magnitude of the current is sufficient to exceed greatly the limits of the magnetization curve of the saturable core; and therefore for the major part of the cycle the reactor has no inductance. During the brief period while the current is passing through zero, the magnetic field varies with the current, that is, inductance is present, and a voltage drop appears across the reactor. This voltage drop, occurring over but a brief portion of the cycle,

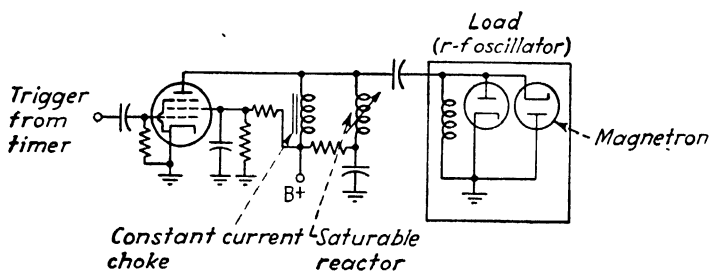



FIG. 336.—Saturable-core reactor driver and modulator.

has a peaked form and represents a pulse of high voltage. Its magnitude is determined by the inductive voltage divider shown. The pulse, superimposed on a negative bias, is applied directly to the grid of the modulator tube. In the circuit shown the peak voltage reaches 1,100 volts, which would drive the modulator 250 volts positive except that grid-current limiting prevents. The pulse width is long judged by usual standards, about 5 μsec. This is one of the disadvantages of the circuit, that is, it is difficult to achieve a pulse shorter than about 0.1 per cent of the pulse interval. Inductive drivers for very short pulses have been designed, however. The 1,100-volt amplitude of the pulse is simply achieved from a 400-volt source.

185. Typical Transmitters and Radiators. Type SCR-268.—Having concluded the general remarks on modulator design, we now return to the principal subject of this chapter, transmitting and radiating equipment. As stated earlier, in this field the general principles of design are most conveniently illustrated by specific examples. In the remainder of this chapter we shall

examine typical equipment[1] operating on 200, 600, 3,000, and 10,000 megacycles.

The first equipment to be described is also one of the earliest designs, the SCR-268 searchlight-control radar put into production in 1940. This transmitter produces 75-kw pulses, 5 μsec long, at a rate of 4,098 per second. The transmitter operates at a frequency of 205 megacycles and employs separate antennas for transmission and reception.

FIG. 337.—Modulator and r-f oscillator of the SCR-268 radar.

The timer and driver of this equipment have already been described briefly (Sec. 180, Chap. VIII). They provide a positive trigger pulse of about 3,500 volts amplitude superimposed on a bias of −3,000 volts. This pulse drives the modulator, a simplified schematic of which is shown in Fig. 337. The circuit is of the hard-tube type, employing series charging and a single capacitor. Resonance is not used, the inductor *L* serving merely to prevent passage of current through the modulator tubes from the power supply. Such current would be subtracted from the current delivered to the oscillator, since the modulator tubes have limited emission.

The modulator tubes are type 304TL (see Table XII); eight tubes are connected to parallel to provide adequate emission. The power supply is a 15-kv d-c unit of the bridge-rectifier type. The storage capacitor is charged through the charging diodes (two type 250TL triodes connected as diodes), and it is dis-

[1] The examples chosen are from among the ground-based equipments developed under the cognizance of the U.S. Army Signal Corps. The techniques are the same as those used in airborne and naval radars, which had not been fully declassified at the time of writing. Details of the latter equipments have since become available in the periodical literature, as listed in the bibliography at the end of Chap. V, page 310.

charged through the oscillator and modulator. The voltage drop through the modulator tubes is about 3,000 volts, leaving 12,000 volts for application to the r-f oscillator. The modulating current has an amplitude of 10 amp. The modulating power applied to the oscillator is accordingly 120 kw, of which about 75 kw is developed as useful r-f power output.

The r-f oscillator employs triodes (type 100TS) in a 16-tube ring circuit (Sec. 152, Chap. VII). This large number of tubes is used because, at the time the design was put into production,

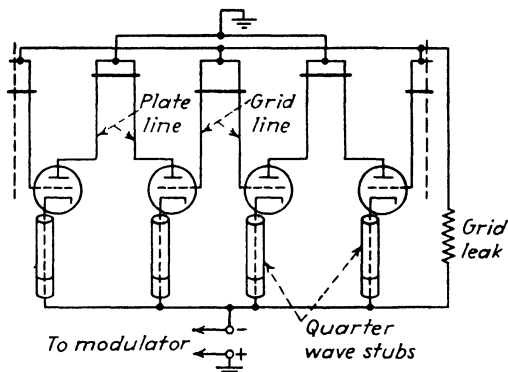



FIG. 338.—Portion of the 16-tube ring oscillator of the SCR-268 radar.

tubes having high emission were not available. A schematic of the transmitter is shown in Fig. 338. The circuit is of the tuned-plate tuned-grid type, with quarter-wave transmission line segments connected to adjacent grids and plates as shown. Concentric quarter-wave lines are used to isolate the r-f filament-heating circuit.

The output of the transmitter is picked up by a loop arranged within the shorting bars of the tuned circuits. These bars form a virtual closed loop since the currents in all bars are in the same direction at any instant. The pickup loop is connected to an open-wire transmission line that conducts the r-f pulses to the radiator. Since separate antennas are used for reception, no duplexing system (t-r box) is required.

The transmitting radiator consists of 16 half-wave dipoles arranged in a mattress, four dipoles high and four wide. Horizontal polarization is used. The reflector consists of similar dipoles spaced a quarter wavelength behind the radiators. The

gain of the radiator is about 50 times that of an isotropic radiator, and this gain is further increased by the gain of the separate receiving antennas, the over-all value being about 100 times The radiated beam has a width at the half-power points of about 10 deg

The dipole elements of the array are fed in phase (the required condition) by cross connections between their outer ends as shown in Fig 339 The transmission line is connected to the termination of the cross connections, which presents a high impedance A matching stub across the open wire is adjusted to match this impedance

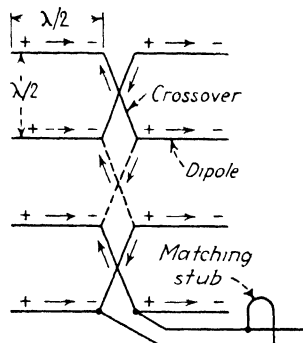



FIG 339 — Mattress and feed of SCR-268 radiator

A view of the SCR-268 radar is shown in Fig 340 The transmitting radiator is the center array, the other two are for reception in the azimuth and elevation angles. The transmitter is mounted on the

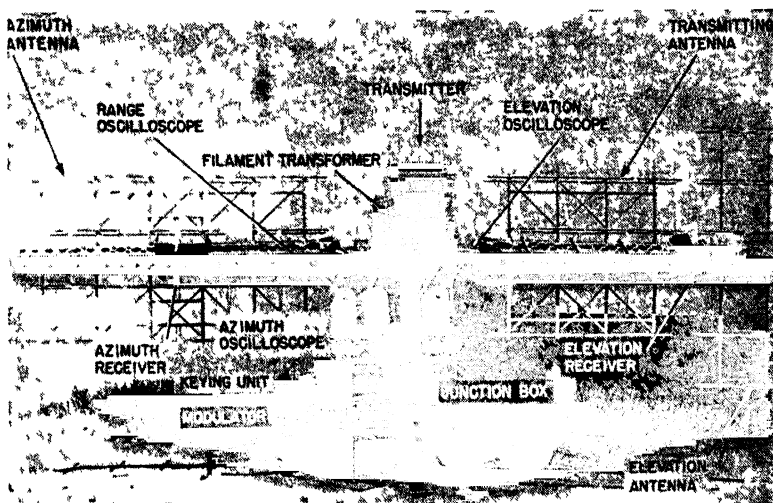



FIG 340 —Components of the SCR-268 radar

top of the pedestal, which rotates as the radar follows its target. A slip-ring system transmits the modulator pulse from the modulator (on the ground beside the pedestal) Details of the remaining

elements (receivers and indicators) are given in the following chapters (Secs. 196 and 209, Chaps. X and XI.

186. The AN/TPS-3 Transmitter and Radiator (500-Megacycle Rotary-gap Equipment).—As a second example we consider the transmitting and radiation equipment of the AN/TPS-3 early-warning radar. This radar operates at 600 megacycles, producing 1.5 μsec pulses at 200 kw peak power and at a rate of 200 per second. The self-synchronous timing system is used. The timing circuits have already been described (Sec. 174, Chap. VIII). The TPS-3 radar represents an excellent design

FIG. 341.—Modulator and r-f oscillator of AN/TPS-3 radar.

for maximum portability combined with long range of detection. Production models of the TPS-3 first appeared in 1944.

The modulator of the TPS-3 radar is shown in Fig. 341. The circuit is recognized as a series-charging version of the d-c resonance type. The primary power source, 110 volts at 400 cps is stepped to about 4,000 volts peak, which is rectified in a voltage-doubling rectifier which produces direct current at 8,000 volts. This voltage charges the pulse-forming line, a constant-k filter of four sections. The total shunt capacitance of this line resonates with the charging choke L at one-half the pulse rate (100 cps). As a consequence of this resonance, the voltage across the pulse line reaches nearly twice the power source voltage (16,000 volts) at the peak of the first positive half cycle. At this instant the rotary spark gap discharges the pulse-forming line through a transmission line and pulse

transformer connected to the r-f oscillator The pulse line thereafter recharges along the resonance characteristic and the cycle repeats

The pulse line is designed to have a characteristic impedance ($Z_0 = \sqrt{L/C}$) of 50 ohms, to match the 50-ohm coaxial cable The delay is 0 75 μsec The rotary gap has three rotary electrodes and one stator electrode; it is driven at 4,000 rpm directly on the shaft of the primary power alternator The discharge

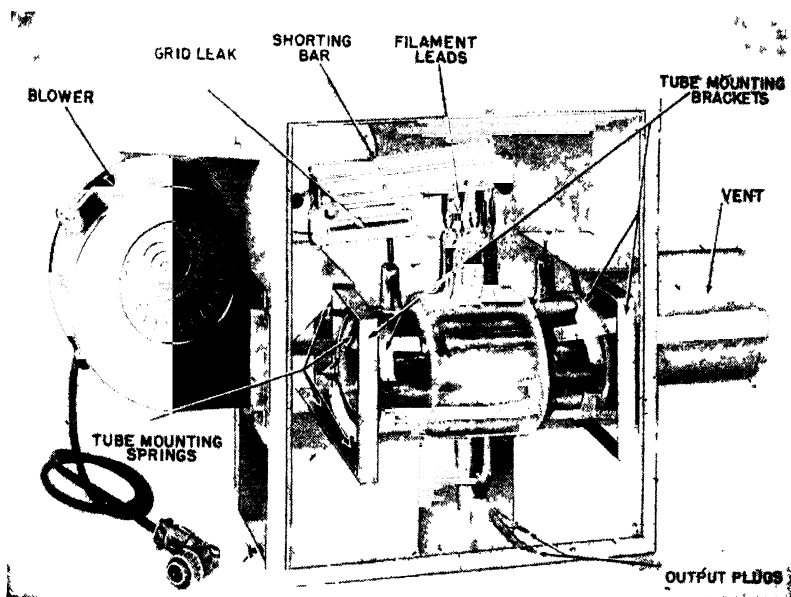



FIG 342—R-f oscillator (internal-circuit tube) of AN/TPS-3 radar

rate is thus 200 per second, one-half the resonance frequency employed in charging the line The discharge current through the gap during the pulse is 150 amp

The pulse-forming line can develop a pulse of only one-half the applied voltage (8,000 volts) This voltage, passed through coaxial cable to the pulse transformer, is stepped up by the transformer to 24 kv The peak power available at the secondary of the transformer is 50 amp at 24 kv, or 1,200 kw, peak Of this input power only 200 kw is developed as useful r-f power output, due to the low efficiency of triodes operating at 600 megacycles The pulse transformer secondary has a double winding through which the filament heating current is passed to the oscillator.

A small auxiliary winding provides the trigger pulse for the timing circuits, as described in Sec. 174.

The r-f oscillator (Fig 342) of the TPS-3 radar is an internal-circuit tube (type VT-158) containing four triode sections and plate and grid resonant lines within the vacuum envelope. The circuit is a push-pull-parallel, tuned-grid oscillator The filaments are of thoriated tungsten; they consume 400 watts heating power and provide a total of 50 amp emission, or 12 5 amp per section.

FIG 343 —Radiator of AN/TPS-3 radar

The modulating pulse (−24,000 volts) is applied to the filaments through a quarter-wave line whose shorting bar is adjusted to prevent loss of radio frequency through the filament heating circuit The grids of the four triodes are fastened to a closed loop conductor that is in effect two quarter-wave sections At one end of the loop, where the r-f voltage is zero, a lead is taken for bias purposes The bias is developed by the grid-leak self-bias method

Power is removed from the oscillator through two leads connected at the center of the plate loop (at the maximum voltage points) These leads are connected directly to the outgoing transmission line, which is held at d-c ground potential by shorted quarter-wave stub grounded at its outer (shorted) end.

Transit time difficulties are minimized in the VT-158 tube not by close spacing but by the use of high anode voltage. Consequently the tube will not oscillate at under 5,000 volts anode voltage, and its efficiency even at 24 kv is under 30 per cent.

The radiator of the TPS-3 (Fig 343) consists of a dipole feed and a paraboloidal reflector. To minimize weight and wind

resistance, the reflector is made of wire netting. The radiator proper, located at the focus of the paraboloid, takes two forms. The simpler of the two (Fig. 344*a*) is a single dipole surmounted by a dipole reflector spaced one quarter wave in front of it. Forward radiation from the radiating dipole is thus prevented (cf. Sec. 97). The beam produced by this combination is about 10 deg wide at the half-power points.

One of the shortcomings of the radiating system is its vertical coverage diagram, shown in Fig. 345. When the radiator has its normal height above terrain of 24 ft, the lowest lobe takes

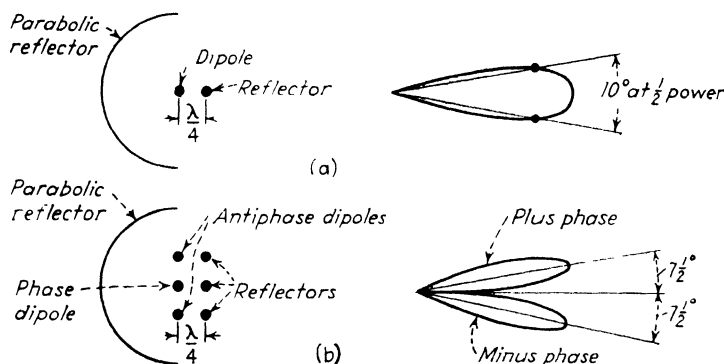



FIG. 344.—Dipole feeds and corresponding beams of AN/TPS-3 radar.

off at an angle about 1 deg above the horizon. Successively higher lobes are spaced at approximately 2-deg intervals. The gaps in the coverage diagram offer an opportunity for intruding enemy aircraft to escape detection. To provide against this contingency, means are provided to shift the lobes into the positions normally occupied by the gaps. The beam shifting is accomplished electrically by shifting the phase among elements of a triple dipole feed (Fig. 344*b*). When the center radiating dipole (so-called "phase dipole") is fed alone, the normal vertical diagram is produced. If, on the other hand, only the two outer dipoles (antiphase dipoles) are fed, the main lobe may be split and caused to occupy two positions 7.5 deg above and below the normal axis. In this case, the outer dipoles are fed 180 deg out of phase by connecting the left half of the upper dipole and the right half of the lower dipole to the center conductor of the coaxial feed line and the remaining portions

to the outer conductor. A magnetic switch, actuated by the operator, connects either the phase dipole or the antiphase dipoles, and thereby shifts the beam as the occasion demands. The relative positions and magnitudes of the vertical coverage lobes in the two cases are shown in Fig. 345. To avoid standing

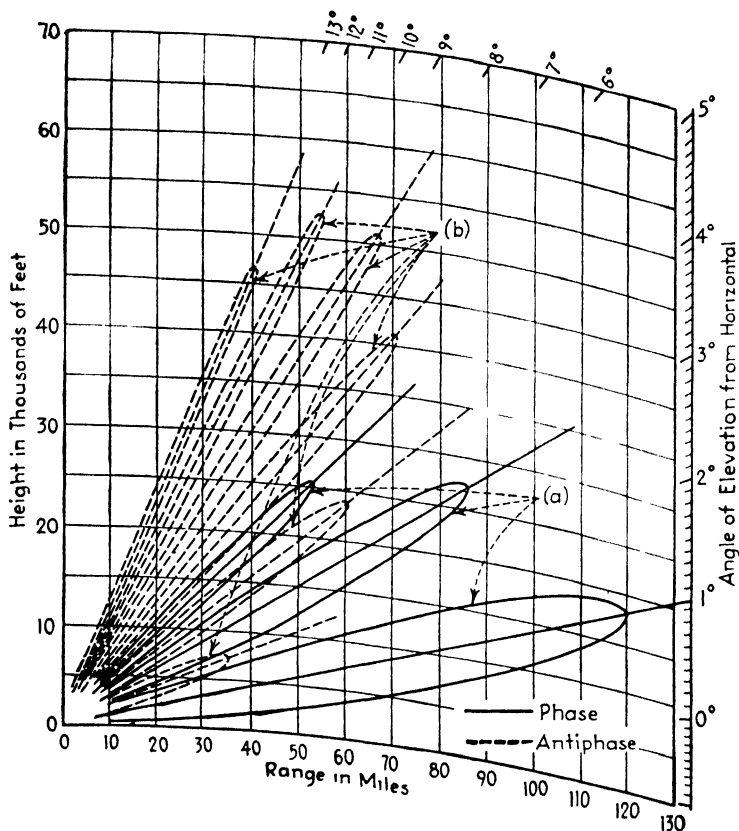



FIG. 345.—Vertical coverage diagram of AN/TPS-3 radar: (*a*) with single-dipole feed, (*b*) with double-dipole feed (compare Fig. 344).

waves, the radiating elements are matched to the stub-supported coaxial 50-ohm feed line by quarter-wave transformers.

Two other elements of interest in the r-f transmission line system are the t-r box and the rotary joint. The t-r system (Fig. 346) includes an anti-t-r element, an open spark gap at the end of a quarter-wave stub, shown at the right in the figure. This gap breaks down during the transmitted pulse, but has no

other effect since the short circuit formed by the arc is transformed to an open circuit at the transmission line connection. The t-r box is a tuned cavity containing a low-pressure t-r tube. The cavity is tuned by a sliding dielectric plug. The transmitted signal, taken from the line at point A, is transformed to the cavity by a coupling loop, exciting the cavity and causing the t-r tube to break down. This detunes the cavity and prevents the transmitted signal from passing through the output loop of the cavity to the receiver.

FIG. 346.—T-r (duplexing) system of AN/TPS-3 radar.

During the intervals between transmitted pulses both spark gaps are inactive. Consequently the received echo pulses are passed through the t-r cavity to the receiver with but small loss. The transmitter oscillator circuits would normally absorb a large portion of the received signal, were it not for the presence of the anti-t-r gap. The open circuit at the end of the quarter-wave stub, created at the inactive gap, is transformed to a closed circuit at the transmission line. This short circuit is transformed, through the quarter-wave spacing, to another open circuit at the junction of the t-r box, which effectually removes from the circuit the transmission system to the right of that point.

The rotary joint, which permits turning the radiator in azimuth (Fig. 347), consists of a capacitance joint with a quarter-wave choke section in the outer conductor, and a self-contained quarter-wave choke within the inner conductor, formed about

the central inner wire which serves as the shaft on which the assembly turns.

The transmitter is matched to the transmission line system through a matching section of two tuning stubs, spaced 3/8 wavelength. Adjustment of these stubs permits matching the impedance presented by the transmitter provided it is greater than one-half the characteristic impedance of the transmission line.

FIG. 347.—Coaxial rotating joint at base of AN/TPS-3 radiator.

187. The SCR-584 Transmitter and Radiator. (3,000-megacycle Gunfire-control Equipment.)—The third example of transmitting and radiating equipment is that of the type SCR-584 radar, the timing circuits of which have already been described in Sec. 177, Chap. VIII. This radar produces 300-kw pulses at approximately 3,000 megacycles (10-cm wavelength). The pulses are 0.8 μsec long and are transmitted at a rate of 1,707 per second. This radar is intended for automatic gunfire control and must provide extremely high accuracy in range and angle measurements. The stringent requirements on pulse timing, duration, and shape are satisfied by the use of a master-oscillator timer, a hard-tube driver, and hard-tube modulator. No gas tubes are used. The r-f oscillator is a cavity magnetron.

The driver of the SCR-584 is shown in Fig. 348. The preceding timer stages (Sec. 177) provide sharp negative triggers of about 15-volt peak input, with sharp leading edges, spaced 586 μsec (rate 1,707 pps). These are fed to an asymmetrical multivibrator, the asymmetry of which depends principally on the

disparity in the coupling capacitors. The multivibrator responds by forming a 2-μsec rectangular pulse followed by a 584-μsec interval. The negative pulse appearing on plate 2 of the multivibrator is passed to an inverting power amplifier stage that produces 2-μsec positive pulses of about 350 volts peak amplitude. These pulses are applied through a capacitor to the driver proper.

The driver consists of two stages of double beam tetrodes. The first of these (1 type 3E29, two beam tetrodes in a single envelope) amplifies the 2-μsec pulse, inverts it, and passes it to a pulse transformer. This transformer reinverts the pulse to positive polarity and applies it to the second driver stage (2 type 3E29). This stage inverts the pulse and the negative pulse output is fed back through the pulse-forming delay line (constant-k low-pass filter) to the grid of the first driver stage. The delay through the pulse line is 0.8 μsec. Consequently, 0.8 μsec after the positive pulse is applied to the first driver grid, a much larger negative leading edge is fed back to the same grid, cutting the tube off and terminating the pulse. At the conclusion of another 0.8 μsec the trailing edge of the output pulse passes through the delay line and the grid of the first driver resumes its normal bias. By that time the positive pulse from the pre-

FIG. 348.—Detailed schematic of SCR-584 driver.

ceding stage has ceased, and the driver remains inactive until the succeeding pulse is applied to the first driver grid. This straightforward method of generating a rectangular pulse of

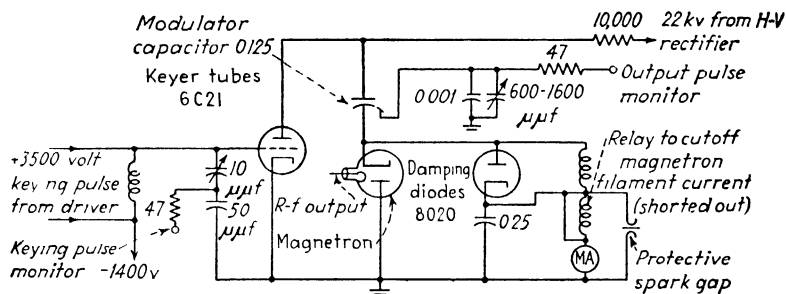



FIG. 349 - Detailed schematic of SCR-584 modulator and r-f oscillator.

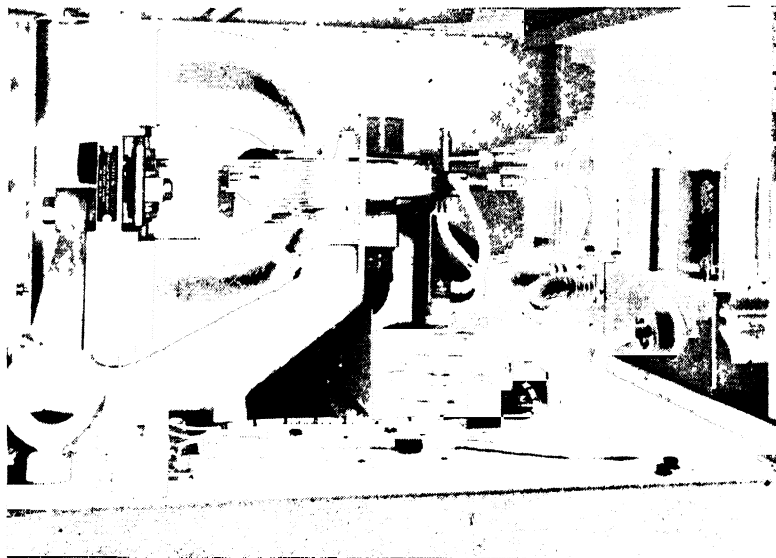



FIG. 350 R-f elements of the SCR-584 transmitter and receiver, showing magnetron and magnet, t-r system, and portion of preamplifier (*right*).

controlled duration depends on the availability of an adequate pulse transformer for coupling the stages in like polarity.

The 0.8-μsec pulse thus formed is coupled from the final driver stage through a second pulse transformer that steps up the pulse to a peak amplitude of 3,500 volts, reverses the polarity, and applies it to the succeeding modulator stage.

The modulator and oscillator stages are shown in Fig. 349. The +3,500-volt driver pulse, superimposed on a 1,400-volt negative bias, is applied to the grids of the modulators, three type 6C21 tubes in parallel (see Table XII). The driver pulse is monitored across a capacitance voltage divider directly at the grids of the modulator tubes.

The driver pulse drives the modulators to full conduction, about 25 amp, which discharges the storage capacitor through the cavity magnetron oscillator. The modulator circuit is of the single-capacitor type, series charged from a fixed d-c supply (resonance is not used). Of the 22 kv across the storage capacitor, about 2 kv is lost in the voltage drop across the modulator tubes, leaving 20 kv at 25 amp for delivery to the oscillator. Of the 500 kw thus applied, about 300 kw is developed by the magnetron (which operates at about 60 per cent efficiency).

FIG. 351.—R-f transmission on line system of SCR-584 radar.

The trailing edge of the pulse is sharpened by the 5-mh inductance between magnetron cathode and ground. The incipient oscillations set up by this inductance are damped out by the damping tubes (three type 8020 diodes). The magnetron is mounted directly on the top of the modulator cabinet, within the poles of a fixed permanent magnet. A view of the magnetron including adjacent r-f elements is shown in Fig. 350. The magnetron operates in accordance with the principles outlined in Chap. IX. The power output is removed by a coaxial fitting, as shown in the figure.

The r-f transmission-line system of the SCR-584 is considerably more complicated than those of the systems previously described, largely because the mount must provide conical and helical

scanning and must therefore move in three coordinates (in azimuth and elevation and about the axis of the paraboloid reflector). The basic transmission line is of the rigid coaxial type, with center conductor supported on broadband quarter-

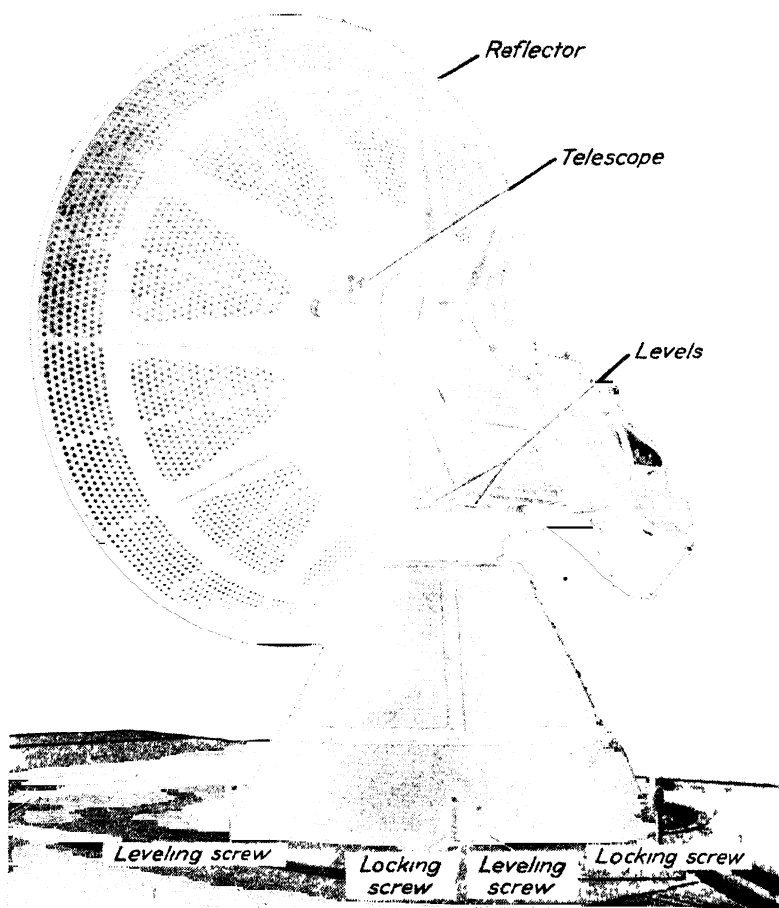



FIG 352 —Radiator of the SCR-584 radar

wave stubs. The r-f system (Fig. 351) starts at the magnetron output and proceeds to the radiator. A tee-junction connects with the t-r box. The main transmission line passes first through a low-speed rotating joint that permits scanning in azimuth, thence through a similar joint for elevation scanning, and finally

through a high-speed joint that permits the dipole radiator to rotate about the paraboloid focus and thus perform conical scanning. The latter joint is in two parts at either end of the dipole motor shaft.

The t-r box (Fig. 353) is of the cavity type, enclosing a type 721-A t-r tube. Tuning is accomplished by two symmetrical screw plugs in either side of the cavity. The t-r box ionizes within a few hundredths of a microsecond of receiving the transmitted pulse and deionizes in about 1 μsec. The attenuation of the transmitted pulse is about 70 db, that is, about 30 mw of power is passed to the receiver. This is sufficient to produce a strong indication, but not sufficient to injure the crystal mixer of the receiver.

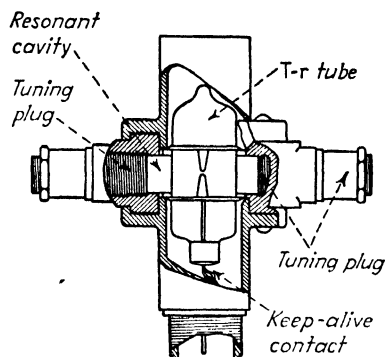



FIG. 353.—T-r cavity of the SCR-584 radar.

The radiating system is a 6-ft paraboloid of perforated metal with a rotating dipole feed. A general view of the radiator and associated rotating joints is shown in Fig. 354. The entire transmission line system is pressure tight and is filled with dry air at 5 lb per sq in. This pressure prevents absorption of moisture that would increase the attenuation of the transmission system. The dipole radiator (Fig. 355) is matched to the coaxial line by a quarter-wave transformer section that converts the unbalanced feed of the transmission line to the balanced feed required by the dipole. This assembly is of the broadband type. The dipole itself is slightly asymmetrical with respect to the axis of rotation, that is, one dipole element is slightly longer than the other. This displaces the center of radiation slightly from the focus of the paraboloid and causes the beam to be radiated off the axis by about 1.25 deg. When the dipole is spun about the reflector axis, the beam thus traces out a cone whose vertex angle is 2.5 deg, as required for conical scanning. In front of the dipole is a flat metal plate that prevents forward radiation. The position of this plate is adjusted to secure an impedance match between radiator and feed line. The trans-

mission line is shorted at the far end, reflecting the signal back to the dipole in proper phase to reinforce the radiation. The whole radiator assembly is covered with a plastic housing that contains the air pressure previously mentioned.

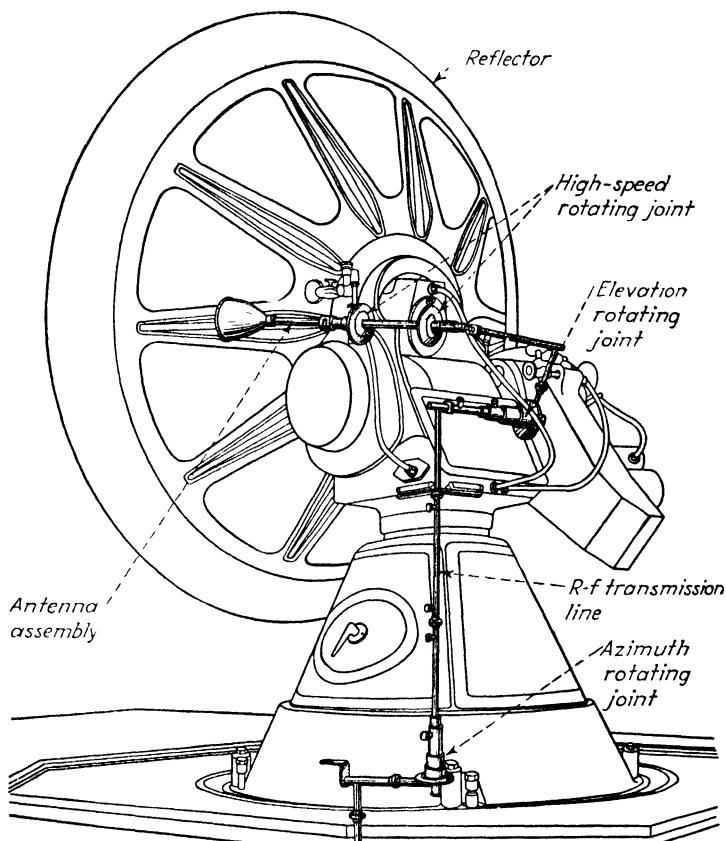



FIG. 354.—R-f transmission-line system, and rotating joints and radiator of SCR-584 radar.

During conical scanning the dipole is spun at 1,800 rpm (30 rps) by a spinner motor, the shaft of which contains the dipole feed line. In consequence of the conical scanning (Sec. 11, Chap. I) the received echo signals have constant amplitude (except for the effects of propagation variations such as fading) when the target is precisely on the axis of the paraboloid reflector. If the target departs from this axis, the echo signals are modulated

in amplitude, at 30 cps, with an approximately sinusoidal envelope. The amplitude of the envelope depends on the extent by which the target departs from the axis of the reflector, and its phase indicates the angular components of the departure in

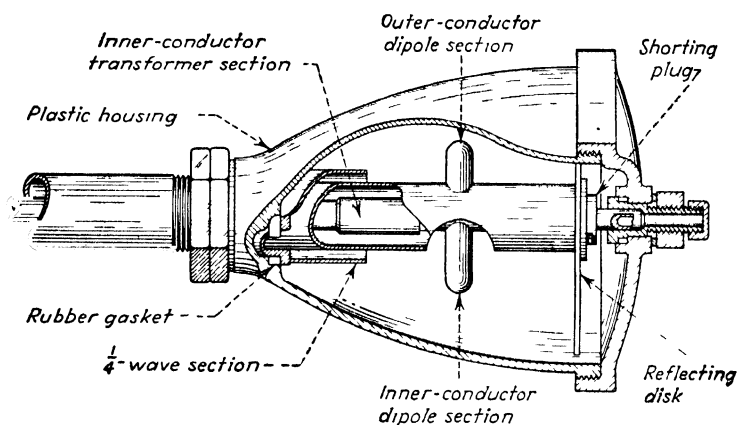



FIG. 355 Rotating-dipole feed of SCR-584 radar

azimuth and elevation. The phase is determined by comparison with a sinusoid generated at 30 cps by a reference generator mounted on the spinner shaft The details of this scanning method and its associated indication and control mechanisms are given in Chap. XI

188. The AN/MPG-1 Transmitter and Radiator System.—As a fourth and final example of transmitting equipment we consider the AN/MPG-1 radar This equipment was designed for harbor and coastal surveillance of, and control of coastal gunfire against, marine targets It differs from the preceding examples in its high operating frequency, 10,000 megacycles (3 cm wavelength), and its use of waveguides in the r-f transmission system.

The MPG-1 radar generates 10,000-megacycle pulses at 35 to 40 kw peak power. Two pulse lengths and rates are provided. A rate of 1,024 pps with a 1-μsec pulse is used for ppi (plan-position indicator) observation where maximum detection range is essential but accuracy is of secondary importance A short pulse, 0.25 μsec, at a pulse rate of 4,094 per second is employed with a type B indicating system when the radar is used for gunfire control

The driver of the MPG-1 is shown in Fig. 356. The timer equipment delivers a 2-μsec trigger with a sharp leading edge, at either of the two rates, 1,024 or 4,094 pps. The trigger pulse is first passed through a differentiating network and an amplifier stage. The output of this stage is fed by a pulse transformer to a cathode follower that aids in matching the impedance of the following pulse-forming line.

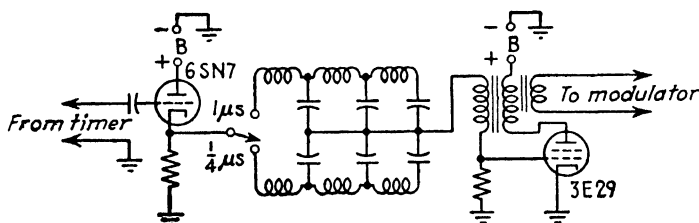



FIG. 356.—Simplified schematic of AN/MPG-1 driver.

The sharp positive pulse from the cathode follower actuates the driver proper, a line-controlled blocking oscillator employing a double-beam tetrode (3E29). Two pulse-forming lines are available for producing 1-μsec or 0.25-μsec pulses. The switch which selects the pulse line is ganged to the switch which controls the pulse rate furnished by the timer equipment. The blocking oscillator operates in the manner described in Sec. 184. The positive trigger from the cathode follower initiates the action of the blocking oscillator, whose grid is driven rapidly positive by feedback through the oscillation transformer. The fed-back voltage charges the pulse-forming line. After two traversals of the line, the line suddenly becomes fully charged. This abruptly reduces the grid potential and the change is amplified cumulatively by the feedback, terminating the pulse. A tertiary winding on the feedback transformer conveys the pulse in positive polarity to the modulator stage, at an amplitude of approximately 1,000 volts.

The modulator and r-f stages of the MPG-1 transmitter are shown in Fig. 357. The modulators are biased at about −750 volts between pulses and are driven about 250 volts positive when the driver pulse appears. The resulting conduction current through the two 5D21 modulator tubes (see Table XII) is about 10 amp, and the anode voltage drop through the tubes is about 1,500 volts. The storage capacitor is charged to 12.5 kv by a

conventional d-c supply. The net modulator pulse is accordingly 10 amp at 11 kv, or about 110 kw peak power.

The modulating pulse is not applied directly to the r-f oscillator but is first passed through two pulse transformers (ratios of 4:1 and 1:4, respectively) and an intervening length of 50-ohm coaxial cable. The second of these pulse transformers has a

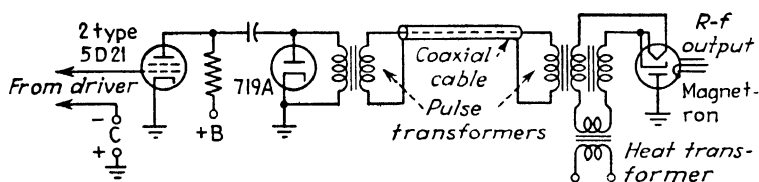



FIG. 357.—Modulator and r-f oscillator, with intermediate coaxial cable, of AN/MPG-1 radar.

double-winding secondary through which the cathode-heating current of the magnetron is supplied. A diode clipping tube is used to limit base-line oscillations. This tube (type 719A) is placed in the modulator, that is, before the first output pulse transformer, but its effect is the same as if connected directly to the magnetron.

The magnetron is a 3-cm oscillator fitted in an external permanent magnet of 5,400 gauss. The magnetron has an efficiency of about 35 per cent, producing about 38 kw peak power output from the 110-kw modulating pulse. The tube has a waveguide output connection, formed by a probe that connects with an internal loop and extends across the waveguide cross section.

The output of the magnetron is conveyed through a rectangular waveguide of standard cross section for 3-cm waves. The elements of the r-f transmission system are shown in Fig. 358. Transmission occurs in the $TE_{1,0}$ mode. The first element after the magnetron coupling is the squeeze box, an impedance-matching device that operates by changing the longer dimension of the waveguide cross section. The waveguide is compressed by an external vise-like structure. Longitudinal slots in the wide face of the guide permit the compression. This device is adjusted, to vary the load on the magnetron, until the most stable operating point is reached. Usually, the point sought is that of maximum frequency stability, rather than maximum power output.

Following the squeeze box is the duplex system, which consists of t-r and anti-t-r cavities, employing 1B24 and 724-A t-r tubes. These operate in the conventional manner (cf. Sec. 158, Chap. VII). Also shown in the diagram are elements of the receiver, two mixer crystals and the local oscillator. These are located immediately adjacent to the t-r box to minimize losses. Their functions are described in the section on the MPG-1 receiver (Sec. 197, Chap. X).

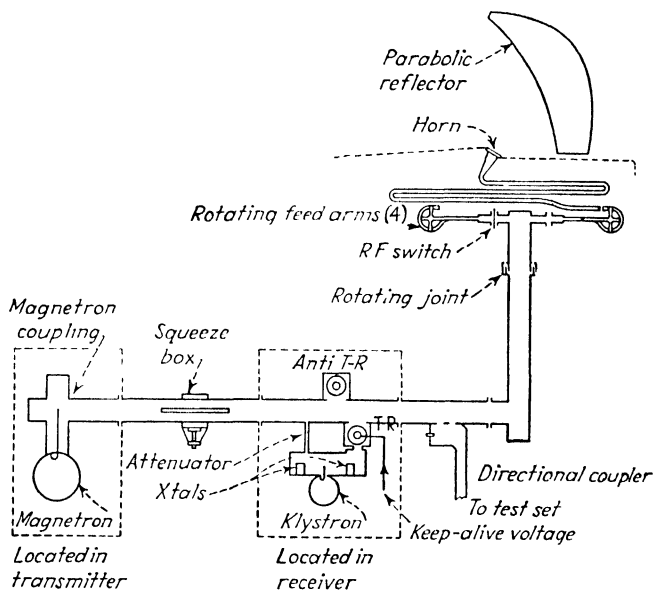



FIG. 358.—Transmission (waveguide) system and r-f elements of AN/MPG-1 radar.

To permit continuous monitoring of the transmitter output without loss of echo signals, a directional coupler (cf. Sec. 219, Chap. XII) is next connected to the waveguide system. This abstracts a small portion of the outgoing energy but has no effect on incoming energy. Immediately following is a butt-type flexible joint in the waveguide, which takes up vibration of the radiator.

The rectangular waveguide, operating in the $TE_{1,0}$ mode, is then coupled to a circular waveguide, operating in the axially symmetrical $TM_{0,1}$ mode to permit rotation in azimuth. The mode conversion is accomplished by a matching diaphragm and

resonant ring near the junction of the two types of waveguide. The details of this coupling section and the succeeding rotary joint and r-f switch are shown in Fig. 359 An external quarter-wave choke is attached at the rotary joint. Surmounting this joint is an r-f switch that feeds the r-f energy successively to one of four waveguide arms in the radiating system. Each of these arms is of the rectangular type, so a second transformation must take place from circular $TM_{0,1}$ to rectangular $TE_{1,0}$ guide Two diaphragms and a resonant ring form the mode-matching

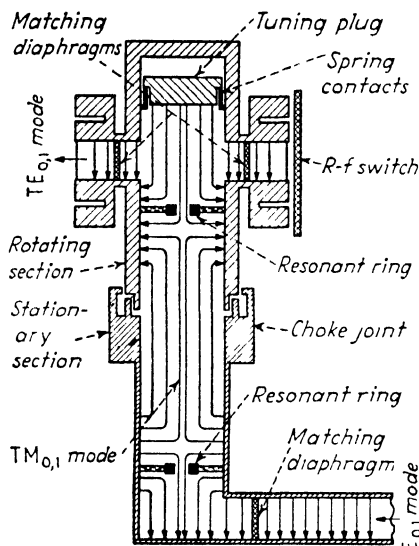



FIG. 359.—Rotating waveguide joint and r-f switch of AN/MPG-1 radar

elements. The rotating assembly is tuned as a whole by the tuning plug at the top.

The radiator proper has already been described in Chap. VII, Sec. 162, to which the reader should refer. The waveguide feed arms move past the throat of a sectoral horn, causing the phase of the energy emerging from the mouth of the horn to vary from one side of the horn to the other. The phase variation causes a corresponding swing in the direction of the beam, over a range of 10 deg in 1⁄16 sec. The beam thus scans in azimuth at a rate of about 160 deg per sec. The relative positions of these elements are shown in side view in Fig. 358.

CHAPTER X

RECEIVERS

The radar receiver is in many respects the most critical element of the system. This is true because it must satisfy many criteria of performance simultaneously and because even minor aberrations in its design or adjustment seriously limit the ability of the system to detect distant targets. Before considering specific receiver circuits and structures, we shall review briefly the functions of the receiver and the technical requirements it must meet.

189. Receiver Functions and Components.—The primary function of the receiver is *detection* of very weak signals, reflected from distant targets and delivered to the receiver input terminals at a power level of a few tenths of a micromicrowatt. By detection is meant discrimination of the signal from the noise. If the only noise present were that picked up by the antenna from the space surrounding the target, the performance of the system could not be improved by the receiver. But other sources of noise, within the receiver, are always present, and these noise sources determine the minimum reflected signal to which the receiver can offer a discernible response. Removal or reduction of these noise sources is the primary goal in receiver design.

The second function of the receiver is *amplification* of the detected signal so that it can actuate an indicator. A very considerable degree of amplification is required. The usual cathode-ray indicator tube requires a control voltage of the order of tens of volts across an impedance of the order of thousands of ohms, or power of the order of tenths of a watt. The minimum discernible signal is of the order of 10^{-13} watt. The power gain required is thus about 10^{12} times, or 120 db. This corresponds to a voltage amplification of a million times, that is, amplification from microvolts to volts.

The amplification requirement is not markedly different, in magnitude, from that met by many other types of receivers, particularly those designed for short-wave communication. But

the signal amplified is different in one important respect, the wide spectrum it occupies. The high gain amplifier must be a wideband amplifier as well. This implies low gain per stage and consequently a large number of amplifier stages. Proper choice of bandwidth is required to secure optimum performance, particularly to minimize the effects of noise.

The choice of the amplification frequency (radio, intermediate, or video) and the relative proportion of gain in each, is open to the designer. For reasons discussed in Sec. 191, the choice usually adopted is as follows: Sufficient gain at radio frequencies (if it can be obtained) is provided to raise the signal above the noise level of the succeeding mixer and i-f (intermediate-frequency) amplifier stages. The main portion of the amplification is carried out at intermediate frequencies, raising the signal level to the order of 1 to 10 volts, so that a high-level linear detector may be used. This is followed by sufficient video amplification to raise the peak signal level to the level demanded by the control electrode or deflection plates of the indicator tube, usually between 10 and 100 volts.

The third function provided by the receiver is *tuning adjustment.* The receiver must display maximum sensitivity at the radio frequency of the transmitter. This is nominally a fixed quantity, but actually it varies over a range comparable with the acceptance band of the receiver. The primary cause of such variations is "pulling" of the transmitter circuit due to impedance variations, caused by scanning and reflected to the oscillator. These variations occur at rates comparable with the scanning rates and cannot be corrected manually. Two methods may be used to insure proper tuning. One is to set up a wider acceptance band than is required by the pulse spectrum and to tolerate the additional noise thereby introduced. The preferred method is to provide automatic-frequency-control circuits. Other frequency variations arise when transmitter tubes are changed, or when the r-f (radio-frequency) system is adjusted to reduce standing waves, etc. These can be corrected by manual tuning adjustments.

A fourth requirement in the receiver is *rapid recovery* from the effects of strong signals. The protection afforded by the t-r (transmit-receive) box is sufficient to prevent injury to the receiver input circuit, but the residual signal admitted by the t-r system

is very strong compared to the reflected signals from small nearby targets. In this case, and in similar cases of strong echoes closely followed by weaker ones, rapid recovery is essential. Throughout the receiver circuits, particularly in i-f and v-f (video-frequency) amplifier coupling connections, care must be taken to avoid accumulation of charge on capacitors that may block the receiver.

In certain radar systems, continuous control of the i-f amplifier gain is provided throughout the pulse interval, the gain increasing as a function of time. This gain control tends to equalize the amplitude of all echo signals, irrespective of distance.

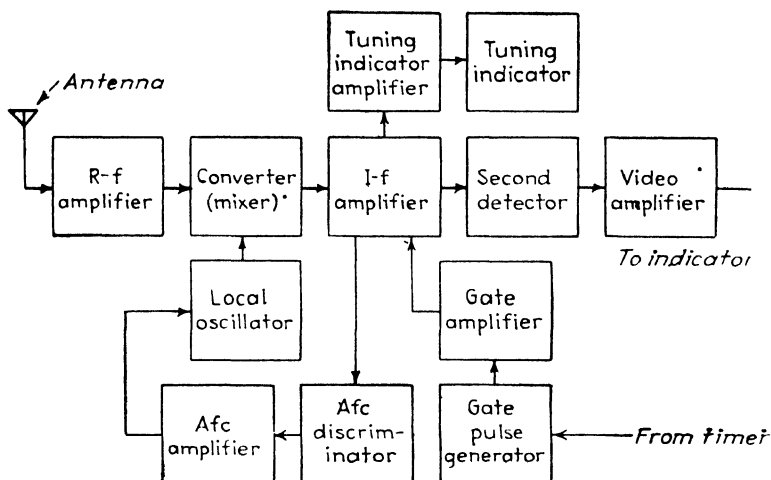



FIG. 360.—Essential elements of a radar receiver.

A fifth and final requirement in certain radar receivers is *selective reception* of echo pulses. In automatic gunfire-control radars, for example, the view of the automatic following circuit must be restricted to a single target, from among many which may lie along the axis of the beam. A similar system is required in receivers that scan the range coordinate while searching for targets in motion (pulse-modulated doppler system). The selection of echoes at particular ranges is accomplished by "gating" the receiver, that is, by activating the i-f amplifier only during a brief period, within which the desired echo is received.

The components that make up a receiver have for the most part been described in previous chapters. They include the r-f

amplifier, the mixer (frequency converter), local oscillator, i-f amplifier, second detector, video amplifier, afc (automatic-frequency-control) discriminator and amplifier, tuning indicator, gate pulse generator, and amplifier. These components and their interrelations are shown in the block diagram of a typical receiver (Fig. 360).

190. Noise in Receivers.—Since the noise contributed by receiver circuits is such an important quantity we shall begin our investigation of receiver design with an examination of noise sources and the derivation of the over-all noise figure of a receiver.[1] As we have seen in Chaps. I and II (Secs. 30 and 60), the basis for noise measurements is taken as the irreducible noise, arising externally to the receiver, which appears across the radiation resistance of the receiving antenna. The magnitude of this noise (the available noise power) is

$$P_n = kT\,\Delta f \qquad \text{watts} \tag{482}$$

where k is 1.37×10^{-23} watt per degree kelvin per cps of bandwidth, T is the absolute temperature of the space coupled to the antenna, and Δf is the effective band width in cps of the receiving system. Since the temperature of space is not a definite quantity (it is known to be variable and to have a minimum value not far above zero absolute), a standard of measurement is adopted by taking T as 290°K (room temperature). For rating receivers on a comparative basis, the antenna is removed and a resistance equal to the radiation resistance, held at 290°K, is connected in its place.

The additional noise contributed by the receiver circuits arises in several cascaded networks: the r-f amplifier (if one is used), the frequency converter, and the initial i-f amplifiers. In a well-designed system, the high level i-f amplifiers and video amplifiers contribute such a small amount of noise that it can be neglected. The cascaded noise sources are indicated symbolically as networks 1, 2, and 3 in Fig. 361. Each is considered to have the same effective bandwidth Δf and the same absolute temperature T, but each has a different gain G_1, G_2, G_3, and a different noise contribution.

[1] FRIIS, H. T., Noise Figures of Radio Receivers, *Proc. I.R.E.*, **32**, (7) 419 (July, 1944).

The noise contribution of each network is represented by a noise figure n_1, n_2, n_3. The symbol n in each case represents the ratio of the noise actually present at the output of the particular network to the noise that would be present if the network were replaced by a loss-less (nondissipative) network.

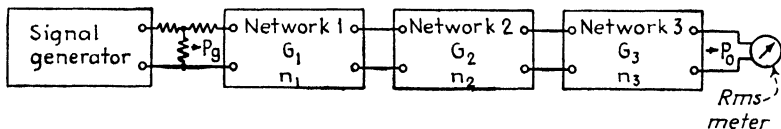



FIG. 361.—Cascaded networks and factors in computation of over-all noise figure.

The noise figure of each network is determined by signal-to-noise ratio measurements. A signal generator of output available power P_g is applied to the input terminals of a particular network (number 1) through an attenuator. The signal-to-noise ratio at the *input* to that network is then

$$r_g = \frac{P_g}{kT\,\Delta f} \tag{483}$$

The signal-to-noise ratio at the output of network number 1 is measured by noting the rms output with the signal generator on, compared with the output with the generator off (but with the attenuator connected). The output signal-to-noise ratio so measured is $r_0 = P_0/N_0$ where P_0 is the available signal power and N_0 the available noise power at the network output. Then the noise figure of network number 1 is

$$n_1 = \frac{r_g}{r_0} = \frac{P_g N_0}{P_0 kT\,\Delta f} \tag{484}$$

But the power gain of the network is $G_1 = P_0/P_g$, so

$$n_1 = \frac{N_0}{G_1 kT\,\Delta f} \tag{485}$$

Here N_0 is the noise actually present at the output of the network and $G_1kT\,\Delta f$ is that portion of the output noise due only to the noise introduced by the signal generator and attenuator. The available noise power P_1 contributed by the network itself is evidently the difference between these two quantities

$$P_1 = N_0 - G_1kT\,\Delta f = (n_1 - 1)G_1kT\,\Delta f \tag{486}$$

If we consider the three networks in cascade and wish to determine the over-all noise figure n, from the component noise figures n_1, n_2, n_3, we proceed as above. The noise output is equal to the noise from the signal generator $kT\,\Delta f$, multiplied by the gain of the whole system $G_1G_2G_3$ multiplied by the noise figure n of the whole system

$$N_0 = nkT\,\Delta f\,G_1G_2G_3 \tag{487}$$

The noise output is also given by the noise input $kT\,\Delta f$, multiplied by the over-all gain, plus the individual noise contributions given by Eq. (486). Eliminating between these two expressions for output noise we obtain the over-all noise figure as

$$n = n_1 + \frac{(n_2 - 1)}{G_1} + \frac{(n_3 - 1)}{G_1G_2} \tag{488}$$

This useful expression shows the relative effect of the noise figures of the component networks. The noise figure of the first network is evidently the most important since it enters directly in the over-all figure. The second network noise figure is less important, provided that the gain of the first network is greater than one. The noise figure of the third network is still less important, again provided that the gain G_2 also exceeds unity. Evidently the higher the gains present, the closer the over-all noise figure approaches the noise figure of the first network.

In microwave receivers, the first network is generally a diode or crystal mixer stage, which displays a conversion loss. G_1 is then fractional and the over-all noise figure is influenced substantially not only by the noise figure of the mixer stage, but also by the noise figure of the succeeding i-f amplifier stage.

The foregoing analysis has been carried out entirely in terms of power and power ratios, rather than decibels, since decibels are not convenient when powers are added. It is customary to express noise figures, both over-all and of component parts of the receiver, in decibels. These figures must be converted into ratios before applying Eqs. (486) or (488). The over-all noise figure may be expressed as a ratio of "so many times theoretical," that is, the noise in the output of the receiver is that many times as great as it would be if the antenna were the only noise contributor. The decibel notation is also commonly used to express

the over-all noise figure "relative to theoretical as zero db." Typical values of over-all noise figures are given in Table IX, Chap. V, pages 304–309. In receivers on the lower radio frequencies, 100 to 300 megacycles, noise figures of 5 to 10 db are commonly achieved. At microwave frequencies the figure usually exceeds 10 db, with 15 db as a typical value.

Knowledge of the physical nature of noise sources is of importance in attempting to reduce tle noise figure. The first type is the thermal agitation noise present in resistors. The available noise power from such sources is given in Eq. (482). In practical radar receivers this component of noise is a relatively small fraction of the total. The other sources of noise are commonly treated as though they originated in resistors and equivalent resistance values are used in circuits to represent the effect.

Whenever vacuum tubes are used, three additional sources of noise may be present, and all three are generally important in radar receivers. They are: shot-effect noise, grid-interception noise, and induced grid noise. Shot-effect noise arises as a consequence of the uneven emission of electrons from the cathode of the tube. Electrons are emitted in groups at random times, and the sudden addition of current, as each group leaves the cathode, excites the connected circuit throughout its transmission band. Operating tubes with low values of plate current or with plate current limited by space charge reduces the noise, but these conditions are both opposed to high gain so the net effect on noise figure is questionable.

Grid-interception noise arises in tubes that have more than one positive collecting electrode. The electrons in the stream divide among the collectors in a random fashion. Hence the plate current reaching the anode displays small random variations caused by irregularities in the current collected by the other positive electrodes. For this reason pentodes have generally higher noise figures than triodes, but here again the higher gain available from pentodes at intermediate frequencies offsets the advantage of the triode, and pentodes are used almost exclusively in i-f service. At the radio frequencies used in radar (above 200 megacycles), the pentode tube does not display higher gain, and thus the triode (generally in a grounded-grid circuit) is employed.

Induced grid noise occurs when the frequency of operation of

the tube approaches the inverse of the transit time of the electrons. When a shot-effect irregularity occurs in the emission, the corresponding change in electron density is preserved as the electrons pass the grid. Hence during a certain interval there is an uneven density of charge on either side of the grid and a current is induced thereby in the grid circuit. This current, passing through the grid impedance, creates a noise "signal" on the grid that is reproduced in the plate circuit, adding further to the rms noise voltage at that point. Induced grid noise increases in direct proportion to the frequency of operation, but the total noise power from this source is proportional to the bandwidth, as in other types of noise, when the band is narrow.

After tube types and circuits have been chosen for best noise performance, two additional steps are taken to minimize noise in practical receiver design. The first is a deliberate mismatch of the receiver input to the antenna to achieve an improvement

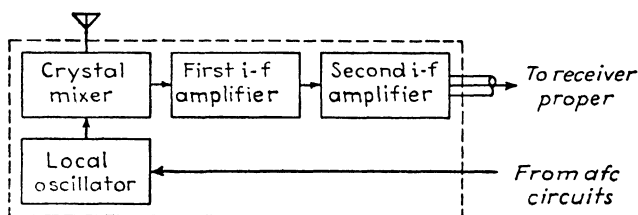



FIG. 362.—Components of preamplifier.

in the over-all noise figure. This adjustment must be performed empirically, with a thermocouple meter in the receiver output. The antenna input circuit is tuned for maximum signal-to-noise ratio rather than for maximum signal. Gains of 1 to 2 db can be achieved in the over-all figure by this method. The second step is to locate the r-f stages of the receiver as close to the t-r box as possible, to minimize losses in the r-f transmission system. The components placed adjacent to the t-r switch are called the "preamplifier" (Fig. 362) and consist of the mixer, local oscillator, and the first two i-f stages. The output level of the second i-f stage is high enough to override noise which might be picked up in the coaxial cable which conducts the signal to the remaining portions of the receiver (i-f stages and video stages).

191. Choice of Amplification Method.—The 120-db gain, more or less, required in a radar receiver might conceivably be obtained

by one type of amplification only, that is, at radio frequencies or at video frequencies. In the first case, applicable only at the lower radar frequencies, a succession of cascaded r-f stages might be employed to reach an output level of 10 volts or more, which, after rectification at high level, could actuate the indicator tube directly. This method is not considered feasible on a number of counts: the gain of r-f amplifiers is low (at microwave frequencies practically nonexistent) and the noise contribution is high. Regenerative feedback from high-level to low-level stages would be very difficult to control, and tuning would have to be applied to all stages.

The alternative, direct conversion from radio frequencies to video frequencies in a detector in the receiver input, followed by a number of video stages, is somewhat more practical. This system has in fact been used in an f-m (frequency-modulation) radar altimeter where the gain requirements are not high and the video signal has a maximum frequency of 10,000 cps. But for general use, the v-f (video-frequency) receiver is impractical. The v-f stage must display l-f gain to avoid introducing a depression of the baseline (undershoot) following each transmitted pulse, which would lower weak near-by echoes below visibility on the indicator. But l-f video gain implies large coupling capacitors between stages, and these, when charged by grid current during strong signals, retain the charge and thus block the receiver for a correspondingly long time. When the initial detector in such a receiver is of the square law type (as are all detectors with very weak signals), the noise figure of the converter is considerably higher than it would be at intermediate frequencies. When high gain and low noise are essential, the video-amplifier receiver is not used.

The system which avoids the foregoing difficulties is the superheterodyne, which introduces an initial change from radio frequency to a lower frequency (intermediate frequency) followed by a high-gain amplification at this frequency. Radio-frequency amplification is used, when it can be obtained, only to mitigate the noise introduced in the mixer stage. The major part of the amplification occurs at the intermediate frequency. The gain of an i-f stage is greater than that of an r-f amplifier, and its recovery time is generally much less than that of a video amplifier. The difficulties with regenerative feedback at intermediate

frequencies are more prominent than at video frequencies, but less so than at radio frequencies. When feedback cannot otherwise be prevented, two frequency conversions and two intermediate frequencies may be used. Intermediate-frequency stages also lend themselves well to continuous or discontinuous automatic and synchronous gain control. The tuning of a superheterodyne is more difficult than that of a receiver employing only a detector and v-f stages, but tuning can be accomplished over a reasonably wide range by adjustment of the local oscillator only, which is susceptible to automatic control.

The choice of the numerical value of the intermediate frequency depends on many factors familiar to designers of conventional superheterodynes. The frequency should be substantially greater than the maximum video frequency to be amplified by the video stages, since filtering the i-f components from the second detector output is then more readily performed. This requires an i-f value of the order of megacycles and preferably not lower than 10 megacycles. Another reason for avoiding too low a value of the intermediate frequency is the application of gate pulses to the i-f stages for discontinuous gain control. If the intermediate frequency lies within the pulse spectrum of the gate pulse, that portion of the pulse spectrum is amplified by the i-f stages. This produces oscillatory tails at the end of the gate pulse, with consequent irregularities in gain. Since the leading edge of the gate pulse has a duration of a few tenths of a microsecond, the intermediate frequency must be above 10 megacycles to avoid this effect. The somewhat larger by-pass capacitors required on the power supply leads of a stage operating at a low intermediate frequency are also a disadvantage when gate pulses are applied, since the response of the amplifier to the gate is then correspondingly sluggish.

Too high a value of intermediate frequency is to be avoided for equally valid reasons. In the first place, the gain per stage tends to drop appreciably above 50 megacycles. Secondly, a high intermediate frequency implies a wide separation between carrier frequency and local oscillator frequency. Since both frequencies are applied to the r-f circuit, it is desirable that the r-f bandwidth extend over both values, although considerable attenuation of the local oscillator signal is permissible and in fact is desirable. The r-f bandwidth at microwave frequencies

(obtainable without resistive loading) is about 50 megacycles, and this gives the order of the maximum permissible intermediate frequency. Since the i-f bandwidth is considerably less than this, difficulties with i-f image frequencies are not serious. Moreover, image response may be tolerated because interfering image signals can be recognized by their lack of synchronism with the indicator sweeps. A third difficulty with high intermediate frequencies is the high degree of misalignment arising from a given casual shift in the shunt capacitance present in the circuit.

The intermediate frequencies used in radar receivers have been standardized, in American and British equipment, at three values: 15 megacycles, 30 megacycles, and 60 megacycles. The 30-megacycle intermediate frequency is most commonly used, particularly in microwave equipment with normal (1-μsec) pulses. The 15-megacycle figure is used in lower frequency radars for amplification of longer pulses. The 60-megacycle intermediate frequency is used in certain gunfire-control radars (particularly those for shipboard use) that employ very narrow pulses.

192. Radio-frequency Components.—The r-f portions of radar receivers have received detailed attention in Chap. VII and hence need not be discussed at length here. The r-f elements are the r-f amplifier (used at low frequencies only), local oscillator, and mixer. These, together with the first two i-f stages, are customarily located close to the t-r box from which they receive the echo signals. Since substantially all the noise in the receiver output arises in these components, special care is taken to design them for maximum gain, to match impedances precisely to minimize reflection losses, and to operate the components at currents and voltages conducive to the minimum noise contribution.

Since all tuning adjustments occur in the r-f section of the receiver, provision must be made in these components for manual and automatic frequency control. Gross changes in tuning, performed manually, may involve many items of equipment, including the t-r box, mixer, and local oscillator. Each of these devices is tuned by adjusting the size of the cavity associated with it. Automatic frequency control is restricted to the local oscillator and is usually limited to a range of 100 megacycles or less, over which the other r-f elements have tolerably uniform

response. The discriminator and afc amplifier that develop the local-oscillator-control voltage are described in Sec. 195 of this chapter.

193. Intermediate-frequency Amplifier Design.—The design of the intermediate amplifier of a radar receiver involves first the choice of the i-f value and bandwidth, second the choice of tubes[1] and circuit configurations, and third the provision of auxiliary gain control and tuning indicator systems. The i-f value is chosen from among the three standard values listed in the preceding section.

The bandwidth must include space for the desired portion of the pulse spectrum plus an allowance for frequency drift of transmitter and/or local oscillator. We recall from Chap. II, Sec. 54, that the amplitude of a rectangular pulse, after transmission, continues to increase as the bandwidth is increased out to the first zero of the pulse spectrum, which occurs at a frequency $1/d$ megacycles, where d is the pulse width in microseconds. When the pulse is transmitted by a carrier, with double sidebands, the same increase in amplitude is noted as the over-all bandwidth, centered about the carrier frequency, is increased to $2/d$ megacycles. Widening the band beyond this limit does not add appreciably to the height of the reproduced pulse, but it does steepen the sides and it generally improves the fidelity of reproduction of the pulse shape.

As the bandwidth is widened, the noise power increases in direct proportion. To determine the bandwidth for maximum amplitude of signal relative to noise, the two functions can be plotted against bandwidth and their ratio taken, as shown in Fig. 363. This figure shows that the maximum signal-to-noise ratio occurs at a bandwidth of about $1.5/d$, or somewhat less than that required for double-sideband transmission out to the first zero of the pulse spectrum. It is evident that the effect of bandwidth on signal-to-noise ratio is not pronounced over the range from $1/d$ to $2/d$. Most practical i-f amplifiers operate within these limits.

The lower value (about $1/d$) was widely used in early designs before it became clear that an additional stage or two of i-f

[1] FORD, G. T., Characteristics of Vacuum Tubes for Radar Intermediate Frequency Amplifiers, *Bell System Tech. J.*, **25**, 385 (July, 1946).

amplification does not impose a severe burden on the system as a whole. The wider band associated with the figure $2/d$ was then more widely adopted, partly to allow extra space for casual frequency shifts, and partly to obtain more faithful reproduction of the pulse.

Wider bands are used in gunfire-control systems. In such systems, high accuracy of measurement of range is essential, so the edges of the pulse (against which the range measurement

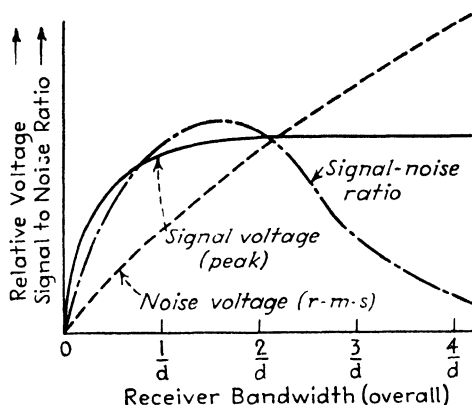



FIG. 363.— Derivation of signal-to-noise ratio as a function of receiver bandwidth.

is made) must be as steep as possible. Signal-to-noise ratio is not a serious matter in many gunfire radars, since the radar is not required to track targets beyond the limited range of the associated guns. In some systems, two i-f amplifiers are used, one with narrow bandwidth for maximum detection range with limited range accuracy, the other with a wider band intended for precise demarcation of near-by targets. The widest bandwidth employed in operational radar during the war was about 10 megacycles, through which 0.25-μsec pulses were transmitted for gunfire control (type AN/MPG-1). This band corresponds to $2.5/d$. A higher ratio would have been preferable for this service, but bands wider than 10 megacycles were difficult to achieve with available tubes. Experimental i-f amplifiers having a bandwidth of 20 megacycles have been built. Another gunfire-control radar, SCR-545, transmits a 1-μsec pulse through an i-f bandwidth of 5 megacycles, corresponding to $5/d$.

When frequency and bandwidth have been established, the

choice of tube and circuit must be made. The tubes must have the highest possible ratio of grid-plate transconductance to shunt-electrode capacitance. This ratio is a figure of merit commonly applied to wideband amplifier tubes. The tubes outstanding in this respect are type 6AC7 and type 6AK5. The 6AC7 tube has a useful g_m (when operated at maximum permissible ratings), or about 10,000 micromhos, and a capacitance sum of 16 $\mu\mu$f, or a merit ratio of 625. The 6AK5 has a g_m under comparable operating conditions of about 6,000 micromhos and a capacitance sum of 6.8 $\mu\mu$f, that is, a merit ratio of 880. The gain of the type 6AK5 is, in equal circumstances, about 20 per cent greater than that of the 6AC7, and this factor has a pronounced effect on the total gain of a multistage amplifier.

The interstage coupling in i-f amplifiers has undergone an evolutionary development. The earliest radar receivers, following television practice, employed mutually coupled (double-tuned) circuits between i-f stages. The bandpass characteristic of the double-tuned circuit has a flatter top and steeper sides than is possible with a single-tuned circuit, and the transfer impedance is higher for a given shunt capacitance. This means that the double-tuned circuit permits higher gain and greater selectivity than the single-tuned circuit. But the double-tuned circuit is more difficult to align and keep in adjustment. In radar work selectivity is not a paramount requirement, and the steep slopes of a highly selective bandpass may in fact be detrimental to pulse transmission. The space, weight, and power cost of one or two additional i-f stages in a radar system amount to little when compared with the total requirements of the equipment. Moreover, simplicity of adjustment in the field, particularly when replacing i-f tubes, is of the utmost importance. These factors point to the desirability of single-tuned circuits between the i-f stages of a radar receiver. This fact was well appreciated by 1942, and most equipment designed since that date has employed single-tuned circuits.

The two methods of coupling are illustrated in Fig. 364. Following the reasoning of Chap. II, Sec. 49, the shunt capacitance is limited to the unavoidable value present in the tubes and wiring. Tuning is accomplished by variation of the inductive elements, that is, by movable metal slugs or powdered-iron cores. When double-tuned circuits are used, the coupling between

circuits must be carefully adjusted to secure maximum gain as well as the desired width and shape of the bandpass characteristic.

The quantitative factors governing the design of single-tuned i-f stages are derived as follows: Consider the single-tuned circuit of Fig. 364*a*. The shunt capacitance C and the inductive element L are tuned to resonance at the intermediate frequency. The circuit is shunted by a resistance R, which is predominantly a resistance shunted across the tuned circuit for the specific

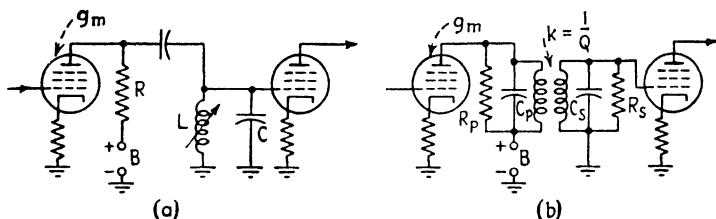



FIG. 364.—Intermediate-frequency amplifier coupling methods: (*a*) single-tuned and (*b*) double-tuned.

purpose of providing the desired bandwidth. Its value includes, also, the resistance of other shunt-circuit resistances and the internal plate resistance of the preceding tube.

The voltage gain of the stage, by analogy with Eq. (437), Chap. VI, is

$$G = g_m Z \tag{489}$$

where g_m is the grid-plate transconductance of the stage and Z is the impedance of the single-tuned coupling circuit at the frequency under consideration. The stage gain is evaluated as a function of frequency from

$$\frac{1}{Z} = \frac{1}{R} + \frac{1}{j\omega L} + j\omega C \tag{490}$$

where $\omega = 2\pi f$ is the angular frequency. If we designate the resonant frequency as $\omega_0 = 2\pi f_0 = 1/\sqrt{LC}$, we can manipulate Eq. (490) to give Z in terms of R as

$$\frac{R}{Z} = \frac{1}{\sqrt{1 + [R\sqrt{C/L}\,(\omega/\omega_0 - \omega_0/\omega)]^2}} \tag{491}$$

A plot of this expression, giving the relative amplitude of Z as a function of frequency, is shown in Fig. 365. We note that the

impedance is a maximum and equal to R at the resonance frequency. Either side of resonance the impedance drops off at a rate depending on the quantity $R/(\sqrt{L/C})$. The larger the value of R, relative to $\sqrt{L/C}$, the narrower the response curve. In particular, the nominal bandwidth of the circuit

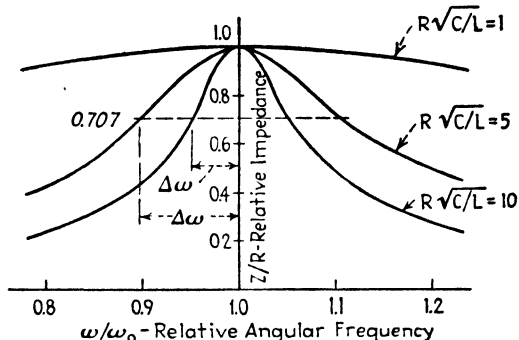



FIG. 365. – Bandpass curves of a single-tuned i-f amplifier.

is the frequency interval $2\Delta\omega$, between the impedance levels 3 db below the maximum, as shown in the figure. This is the so-called "half-power" bandwidth, at which the voltage gain of the stage falls to 70.7 per cent of its value at resonance. If we solve Eq. (490) for ω, when $Z/R = 0.707$, on the assumption that the bandwidth is small compared with the resonant frequency, we obtain the frequencies of the band edges, ω_1 and ω_2, such that

$$2\Delta\omega = \omega_2 - \omega_1 = \omega_0 \frac{\sqrt{L/C}}{R} = \frac{1}{RC} \tag{492}$$

This is the nominal bandwidth of a single stage, within which the stage gain varies between the limits $0.707g_mR$ and g_mR. The expression shows that for maximum bandwidth R and C must be small. The shunt capacitance is accordingly made as small as the circuit structure will allow, and R is chosen to produce the required bandwidth. For a given value of C, as the bandwidth is increased, the value of R must decrease and the gain of the stage must decrease proportionately.

When a number of single-tuned stages are connected in cascade, all tuned to the same resonant frequency, the over-all bandwidth of the amplifier is less than that of a single stage. To

illustrate this effect, consider n identical stages. The over-all bandpass characteristic of the n stages is the product of the characteristic of the single stage, shown in Fig. 366, multiplied by itself n times. The nominal over-all bandwidth of the amplifier is, as before, the frequency interval between the 3 db points of the over-all curve. Then

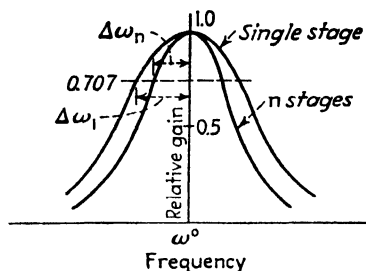



FIG. 366.—Reduction of bandwidth caused by cascading identical stages.

$$\left(\frac{Z}{R}\right)_{\text{over-all}} = \left(\frac{Z}{R}\right)^n = \frac{1}{\{1 + [R\sqrt{C/L}\,(\omega/\omega_0 - \omega_0/\omega)]^2\}^{n/2}} \tag{493}$$

where n is the number of stages. Now, as in the single-stage case, we can find the nominal bandwidth by setting

$$\left(\frac{Z}{R}\right)_{\text{over-all}} = 0.707 = \frac{1}{\sqrt{2}}$$

and solve for the values of ω. This process yields two frequencies ω_1 and ω_2 at the band edges, such that

$$2\Delta\omega_n = \omega_2 - \omega_1 = \frac{\sqrt{2^{1/n} - 1}}{RC} \tag{494}$$

Comparison of this expression with Eq. (492) shows that the bandwidth of the multistage amplifier is equal to that of a single stage multiplied by the factor $\sqrt{2^{1/n} - 1}$, which is a fraction for all values of n higher than 1. Table XIV gives the values of this factor for amplifiers of 1 to 10 cascaded stages.

TABLE XIV.—RELATIVE VALUES OF OVER-ALL BANDWIDTH BETWEEN 3-DB POINTS OF n CASCADED STAGES

Number of stages	Bandwidth	Number of stages	Bandwidth
1	1.00	6	0.35
2	0.64	7	0.32
3	0.51	8	0.30
4	0.44	9	0.28
5	0.39	10	0.27

Since it is not unusual to employ 6 cascaded stages with single-tuned circuits, it follows that the bandwidth of the amplifier is commonly about one-third that of a single stage provided all the stages are tuned to the same frequency. To achieve the desired over-all bandwidth, in other words, it is necessary to design each stage for a bandwidth about three times as great, and the gain per stage is correspondingly low.

A better compromise is obtained when the cascaded stages are tuned not to the same resonant frequency, but to two or three values in successive stages, the values being staggered over the desired over-all band (Fig. 367). This procedure (called

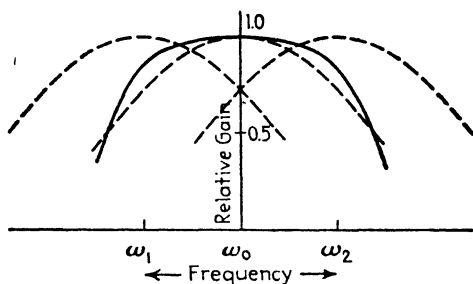



Fig. 367.—Stagger tuning of i-f amplifiers.

"stagger-tuning") makes it possible to obtain an over-all bandwidth that is equal to or slightly greater than the bandwidth of a single stage, at some loss in steepness of the band edges (selectivity). As noted previously, selectivity is not a highly important factor in radar receivers, and the rounded edges of a nonselective characteristic are advantageous in the reproduction of a rectangular pulse. The great advantage of stagger tuning is that each stage may be designed for a narrower bandwidth, and hence higher gain, than is possible if all the stages are to be tuned to the same frequency. The additional gain per stage produces a substantial effect on the over-all gain. The quantitative treatment of stagger-tuned circuits is not so simple as that of identically tuned stages, but yields to straightforward analysis.

The shunt resistance of a single-tuned circuit is commonly placed in shunt with the plate of the preceding tube, rather than in the following grid circuit, which is grounded directly by the inductive element as shown in Fig. 364*a*. This arrangement

has the shortest possible recovery time, since charge accumulated on the coupling capacitor during strong signals is discharged through the resistance in the plate circuit of the preceding tube. The resulting voltage drop has but little effect on the gain and cannot block the stage.

The practical advantage of single-tuned circuits, compared with double-tuned stages, is the ease with which they can be aligned with simple test equipment in the field. Each stage is aligned separately, with a signal generator connected to its grid at the resonant frequency prescribed for that stage, and the inductive tuning element varied until the output of the stage is a maximum. This process is repeated, at the frequencies prescribed, in the other stages. The over-all gain and bandwidth then automatically assume the desired values.

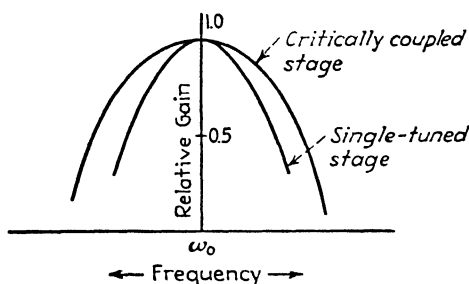



FIG. 368.—Comparison of single- and double-tuned bandpass characteristics

No such simple procedure suffices when double-tuned stages are used. However, the over-all gain for a given bandwidth of double-tuned stages in cascade is greater and this may serve a purpose where the power and space requirements of additional i-f stages cannot be tolerated. A simple quantitative analysis of double-tuned stages (Fig. 364*b*) is possible only if the two circuits (primary and secondary) display the same value of Q and if the mutual coupling between them has the critical value $1/Q$. Then the bandpass curve has the rounded shape shown in Fig. 368, and the maximum gain of the stage is

$$G = \frac{g_m\sqrt{R_pR_s}}{2} \tag{495}$$

where R_p and R_s are the effective resistances shunting the primary and secondary. The maximum gain of n identical stages in

cascade is this quantity raised to the nth power. By a process similar to that employed for single-tuned stages, the over-all bandwidth of n identical double-tuned stages is found to be

$$2\Delta\omega_n = \omega_2 - \omega_1 = \frac{\sqrt[4]{4(2^{1/n} - 1)}}{C_s R_s} \tag{496}$$

where C_s and R_s are the capacitance and resistance of the secondary circuit. For given values of these quantities, the bandwidth of n double-tuned stages is seen to be greater than that of n single-tuned stages in the amount that $\sqrt[4]{4(2^{1/n} - 1)}$ exceeds $\sqrt{2^{1/n} - 1}$. Moreover the primary and secondary stages separate the capacitances associated with each tube, so C_s in Eq. (496) is commonly less than C in Eq. (494). Since the bandwidth is greater, the gain for a given bandwidth can be made greater (by employing values of R_p and R_s larger than R).

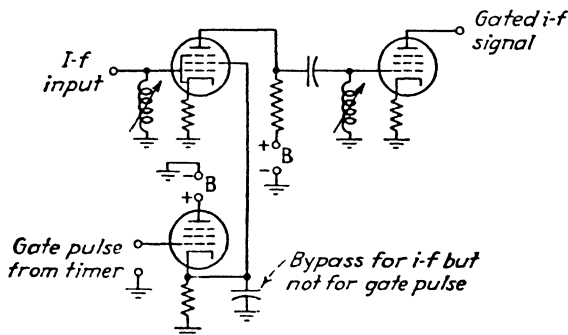



FIG. 369.—Gating of i-f amplifier stage.

The remaining aspect of i-f circuit design relates to discontinuous gain control, that is, gating. This function can be accomplished in a variety of ways. One simple method is to ground the screen grid of an i-f stage through a resistance that serves also as the cathode resistance of a cathode follower stage (Fig. 369). When the positive gating pulse (produced by the timing equipment) is applied to the grid of the cathode follower, its cathode is raised in potential by an equivalent amount. This momentary positive potential is applied to the screen grid of the i-f amplifier stage, which thereupon becomes conducting and amplifies the echo signal applied to it from the preceding stage. The cathode follower stage provides a low resistance

path to ground, thus effectively cutting off the stage between gate pulses.

Positive pulses may also be applied to the control grids of one (or more) i-f stages, the stages being biased to cut off by an external supply between the gate pulses. Intermediate-frequency stages may also be deactivated momentarily by the application of negative gate pulses to the control grid or screen grid. This latter method is employed to promote rapid recovery after the transmitted pulse. Positive gating is used to select desired signals from others or from the background.

The design of gated stages involves not only the amplitude of the positive or negative pulses applied to the gated tubes, which must be properly chosen with respect to bias values and the tube characteristics, but also the *RC* time constant of the circuit through which the gate pulse is applied. The terminals must be by-passed to prevent loss or regeneration of the echo signals in the case of the control grid, or to maintain i-f ground in the case of the screen grid. If the intermediate frequency has a low value the by-pass capacitors must be correspondingly large, and the gate pulse takes a long time to develop across the tube electrode. The response of the receiver i-f amplifier to the gate pulse is then sluggish. The remedy is to use as high an i-f frequency as other considerations will permit, and to apply the gate pulse from a low-impedance source such as a cathode follower stage.

194. Video Detection and Amplification.—The output of the i-f amplifier is conveyed to a second detector and video amplifier, which develops the received pulses at an amplitude adequate for the indicator tubes, while preserving the shape of the pulse. Fortunately, noise is not a consideration at this stage of the reception process, but care must be taken to avoid introducing unwanted transients during video detection and amplification.

The second, or video, detector is almost universally a diode in radar receiver practice. The diode and its load resistor constitute the resistive loading of the last i-f interstage coupling network. Consequently during conduction (while the signal is present) the diode-and-load resistance must be sufficiently low to provide the necessary bandwidth. The resistance of the detector circuit reflected to the preceding i-f interstage coupling is commonly taken as one-half the load resistance value. If the shunt

resistance [cf. R in Eq. (492)] across the interstage coupling is to be, say, 1,000 ohms, the diode load resistance cannot be larger than 2,000 ohms. But with such a low value of load resistance, a small detected output is developed, unless the dynamic internal plate resistance of the diode is of comparable (and preferably smaller) value. This makes a low-resistance diode advantageous in video detection. The 6H6 diode is widely used. Lower resistance diodes are available, and the high-g_m pentodes may be connected as diodes.

In addition to low resistance, the diode should have low anode-cathode capacitance and low electrode capacitance to ground. The anode-cathode capacitance tends to discharge the signal voltage across the load during the nonconducting halves of the i-f cycles. The shunt electrode capacitance to ground contributes to the total shunt capacitance of the stage and thus limits the bandwidth available at a given gain per stage.

The diode detector may be connected in either of two polarities. When the load resistance connects between the detector cathode and ground (so-called "cathode-to-ground" connection), the output pulse has positive polarity with respect to ground. This connection is commonly used when the video stage following the detector must operate cutoff between pulses. The opposite connection, anode-to-ground, produces negative pulses so that the next video amplifier must operate with plate current flowing between pulses. This latter connection, while somewhat wasteful of plate power, has the advantage that a high amplitude pulse is automatically limited by the grid-bias cutoff level of the video amplifier. This self-limiting circuit is particularly useful in compact equipment when only one video stage is used and the output must be applied directly to the control grid of the cathode-ray indicator tube.

The coupling circuit between video detector and the following amplifier must generally be designed as a compromise between faithful reproduction of the pulse and attenuation of the i-f components. While the video stage nominally does not have gain at i-f frequencies, actually appreciable i-f gain may exist. If the i-f components in the detector output are thus amplified, the inferior shielding usually found in the video-amplifier circuit may permit sufficient energy to reach the low-level i-f stages to cause regenerative feedback. The most efficient type of cou-

pling circuit, from the standpoint of separating i-f from v-f components, is the multisection low-pass filter. But this refinement is seldom used. Series peaking coils, designed in conformance with video amplifier practice, are commonly found in detector load circuits. Typical video detector circuits are shown in Fig. 370. The plate circuit triode detector shown is sometimes used but has no particular advantage over the diode form.

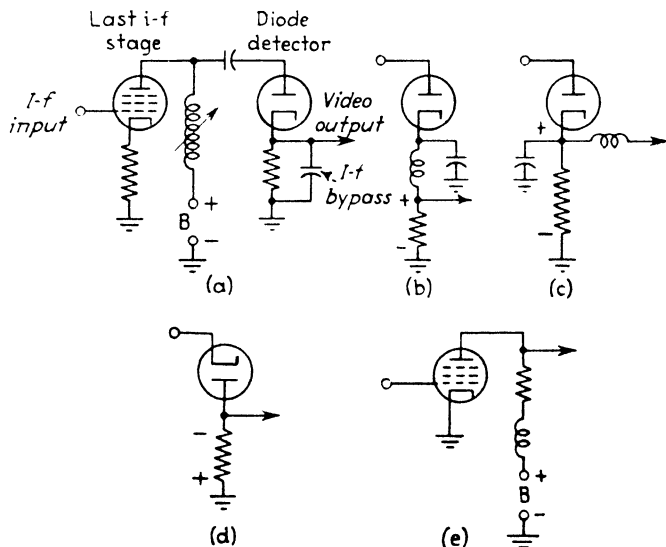



FIG. 370.—Video detectors: (a) diode, uncompensated, (b, c) inductively compensated, (d) connection for negative output, and (e) pentode power detector.

In any event the detector should operate as nearly as possible as a linear device (that is, with peak v-f output linearly related to peak i-f input). This implies that the detector be fed a high-amplitude signal, since the curves of all detectors at low levels tend to have a convex-downward nonlinear shape. This curvature offers higher amplification to strong signals than to weak ones, and this emphasizes the visual effect of background noise when the signal is of smaller magnitude.

Video amplifiers in radar receivers follow the general principles outlined in Secs. 134 through 137, Chap. VI. Inductive peaking (series, shunt, or series-shunt) is commonly used, and the design follows the relations given in Sec. 135. Since the video amplification is generally carried out in but one or two stages, the com-

pensation applied per stage may be liberal without fear of incurring the transient responses of a multistage amplifier. The main consideration of output video amplifiers is the fact that they must develop tens or hundreds of voltage across an impedance that is generally under 5,000 ohms. This implies a

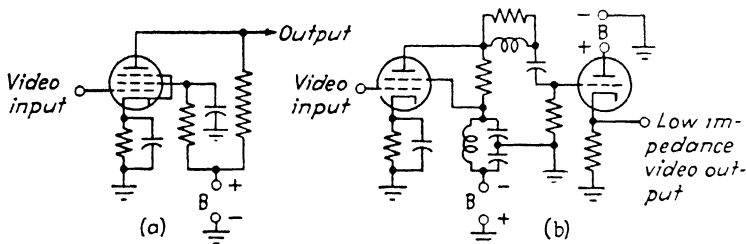



FIG. 371.—Video amplifier connections: (*a*) uncompensated pentode or tetrode with cathode self bias, (*b*) compensated tetrode with cathode-follower output stage.

peak power of the order of watts, with the consequent requirement that the output tube be of the power-amplifier variety. Several forms of beam-power tubes (6AG7, 6V6, and 6L6) are widely used for this purpose. Fig. 371 shows typical video amplifier circuits.

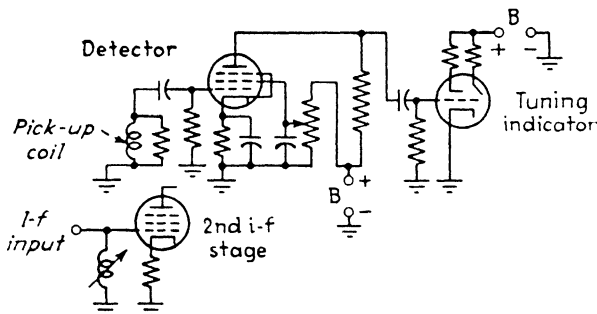



FIG. 372.—Detector and i-f pickup for tuning indication.

195. Tuning Auxiliaries.—Two auxiliary circuits commonly provided in radar receivers are the tuning indicator and the automatic frequency control. The tuning indicator permits adjusting the receiver in the absence of a standard type A indicator. A typical example is shown in Fig. 372. The signal is taken by a small pickup coil from the output of the first i-f stage to a power detector (a high-g_m pentode) that rectifies

the i-f signal and passes the resulting pulse to the "tuning-eye," a type of cathode-ray indicator commonly used in domestic radio receivers. The circuit between power detector and indicator has a long time constant and hence develops a constant voltage proportional to the power detector output. This voltage is so applied to the grid of the indicator tube that the shadow in the fluorescent pattern decreases as the strength of the signal

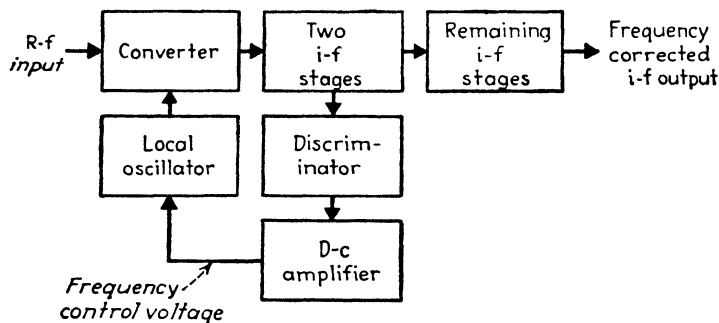



FIG. 373.—Elements of automatic frequency control circuit.

increases. The receiver is tuned for maximum response by adjusting the t-r box, mixer, and local oscillator and by noting the positions of maximum response. The sensitivity of the tuning indicator may be adjusted by varying the screen voltage of the power detector. The power detector may be replaced by an equivalent buffer amplifier and diode detector.

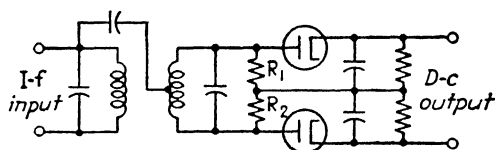



FIG. 374.—Frequency-modulation detector (discriminator).

Automatic frequency control is provided by a discriminator circuit. This circuit develops a direct voltage proportional to the deviation of the intermediate frequency from its assigned value. A typical afc circuit is shown in Fig. 373. The circuit operates on the strong pulses from the transmitter, and the signal is accordingly taken from the low-level i-f stages of the receiver. The essential element of the afc system is the discriminator circuit, shown in Fig. 374. This circuit is widely used as a frequency detector in domestic f-m receivers. The input i-f voltage

is fed to a transformer and voltage is induced in two equal portions on either side of the center-tapped secondary. The primary and secondary are tuned to the assigned intermediate frequency and are loosely coupled, and thus there is little reaction of one circuit on the other. The voltage induced in the secondary is 180 deg out of phase with the primary voltage. This voltage is, however, induced in series with the inductive and capacitive elements of the secondary circuit, so the voltage appearing across the secondary capacitor is 90 deg out of phase with the primary voltage.

To this quadrature voltage is added the primary voltage by the capacitive connection from the primary to the center tap of the secondary. If the signal has the frequency to which the transformer is tuned, the voltages across the load resistors R_1 and R_2 are equal and opposite, and the rectified direct voltages across the diode loads are likewise equal and opposite; therefore there is no direct voltage across the output. But if the intermediate frequency is shifted upward, say, by a variation in transmitter frequency, the induced secondary voltage and the capacitively applied secondary voltage are no longer 90 deg out of phase. The vector addition then produces a higher input voltage to the upper diode and a lower input voltage to the lower diode. Corresponding differences appear in the direct voltages across the output loads and their algebraic sum is then a net positive quantity. Thus an upward shift in intermediate frequency develops a positive direct voltage in the output of the discriminator. Similarly a downward shift produces a negative direct voltage. Over a limited range of frequency shift, the direct voltage generated is closely proportional to the frequency deviation.

The control voltage thus developed by the discriminator may be applied directly to the repeller electrode of the klystron local oscillator, in proper polarity to correct the shift in intermediate frequency. The magnitude of the control voltage (for a given frequency deviation) depends on the magnitude of the i-f signal. The 10-cm reflex klystrons require about 2 to 5 volts to shift frequency 1 megacycle; the 3-cm tubes require about 1 volt per megacycle. If voltages of this order are not obtainable directly from the discriminator output, a d-c amplifier may be interposed between discriminator and local oscillator. Operation from a high-level r-f stage might also serve, but the effect of strong

echo signals must then be taken into account. When these strong echo signals vary in magnitude, as they do during scanning, they may change appreciably the i-f signal applied to the discriminator, and the magnitude of the discriminator output will then vary from causes not associated with frequency deviation. By operating from a low-level I-f stage, the response is limited to the transmitted pulse, which is presumably of constant amplitude.

The output circuit of the discriminator may have a long time constant or a short one, depending on the rapidity with which the frequency correction is to become effective. The simpler case occurs when the frequency correction can occur slowly. Then a time constant longer than the pulse interval is employed and a slowly varying direct voltage appears in response to correspondingly gradual shifts in transmitter or local oscillator frequency. The system is suitable when the shifts are small and when the scanning rate (if scanning causes pulling of the transmitter frequency) is low.

In certain equipments, particularly 3-cm equipments with rapid-scanning radiators, frequency shifts may occur from pulse to pulse, within a small fraction of a second. To correct such shifts it is necessary to employ a very low time constant in the discriminator output. The output voltage then decays between pulses, that is, the discriminator produces an output pulse for each transmitter pulse, the magnitude and sign of the pulse depending on the degree and direction of the frequency deviation from the assigned value.

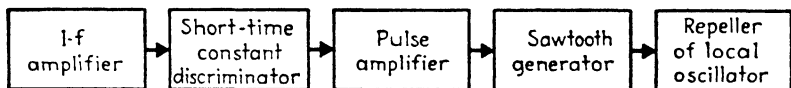



FIG. 375.—Block diagram of sawtooth a-f-c circuit.

It is not feasible to apply these discriminator output pulses directly to the local oscillator, since the oscillator would then shift rapidly, immediately following the transmitted pulse, and would detune the receiver progressively during the reception of the echoes. A circuit which gets around this difficulty (Fig. 375) employs a sawtooth generator which is controlled by the discriminator output pulses. The sawtooth wave swings the local oscillator frequency over a range determined by its amplitude. The intermediate frequency is thus swept through its assigned

value and the discriminator output pulses change sign from negative to positive. These pulses are applied to the grid of the sawtooth wave generator, a gas tube. The negative pulses have no effect on the gas tube, but the positive pulses cause the tube to conduct and thus terminate the sawtooth wave shortly after it has swept the intermediate frequency through its assigned value. In this way the receiver is constrained to be in tune. The amplitude of the sawtooth is adjusted so that the total deviation introduced does not exceed the limit of the acceptance band of the receiver, and the average loss in sensitivity is small. In some cases, this sawtooth sweeping system is set up in two steps, a large voltage swinging the frequency over 100 megacycles to find the correct frequency followed by a second low-amplitude sweep that takes over when the correct frequency is found and disables the high-amplitude sweep. This type of afc system is used in the AN/MPG-1 radar and is described in Sec. 197 of this chapter.

196. Typical P and L Band Radar Receivers. Types SCR-268 and AN/TPS-3.—The first radar receiver we shall examine operates at a low frequency, 205 megacycles. This receiver is employed in the SCR-268 searchlight-control radar, parts of which have been described in previous chapters. This radar represents an early design. The block diagram of the receiver is shown in Fig. 376.

The first r-f stage, employing type 954 acorn pentodes, has two separate channels, which are switched on and off alternately by a rectangular wave applied to the grids of the tubes. The purpose of this switching is to connect the receiver alternately to two transmission lines that are taken from opposite sides of the antenna array. The phase lag across the face of the array causes the direction of maximum sensitivity (axis of the received beam) to shift slightly when the receiver is switched from one transmission line to the other. This technique, known as "lobe switching," has been described in previous chapters. In the SCR-268 receiver, the switching is accomplished by a multivibrator and amplifier that produce a symmetrical rectangular wave at 1,400 cps.

The plate connections of the two r-f tubes are common, and therefore the two lobe-switched signals are combined in a single signal displaying interspersed echo pulses of two different

amplitudes (when the target is off the axis of symmetry). The combined signal is amplified in a second i-f stage and then passed to the mixer. Acorn tubes are used through the r-f system. The local oscillator is a 955 acorn triode, operating at a frequency 19.5 megacycles higher than the carrier frequency. The intermediate frequency is accordingly 19.5 megacycles and is amplified in four i-f stages employing 6SK7 remote-cutoff pentodes. The receiver employs double-tuned interstage coupling and has a nominal bandwidth of about 1 megacycle. The

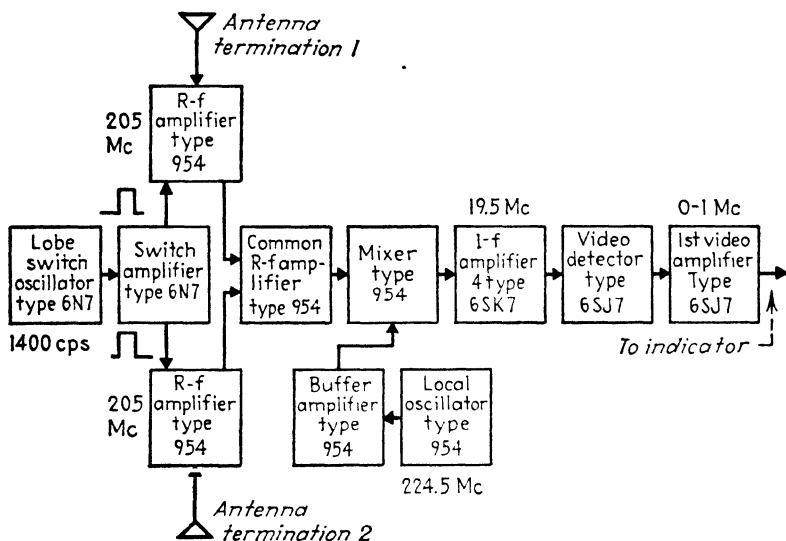



FIG. 376.—Block diagram of SCR-268 receiver.

pulse length is about 5 μsec so the relative bandwidth is about $5/d$. The i-f amplifier has a gain of about 90 db, which, with the 15 db gain in the r-f stages, is sufficient to develop a 1-volt amplitude from the noise across the input. The sensitivity of the i-f stage is controlled by grid bias on the first stage. The noise figure is about 14 db above theoretical.

A sharp cutoff pentode (6SJ7) is used as a second detector of the power type and the pulses are amplified in a video stage using the same type tube. Further video amplification, which takes place in a stage associated with the indicator, is performed in a push-pull beam power stage (type 6L6).

It will be noted that the SCR-268 receiver has few of the refinements of later receivers. Automatic frequency control

is not used, and even the t-r box is omitted since separate receiver and transmitting antennas are used. The sensitivity is adequate, particularly since the radar was designed primarily for controlling searchlights, which have in themselves rather limited range.

The receiver of the AN/TPS-3 radar (Fig. 377) operates at 600 megacycles. It obtains its input directly from the t-r

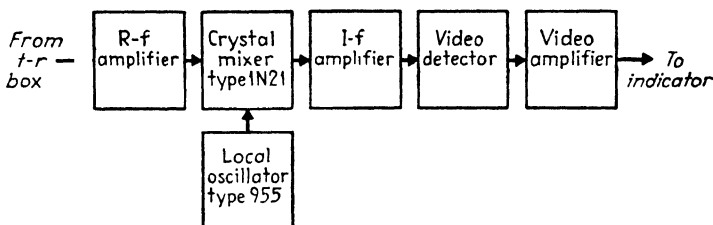



FIG. 377.—Block diagram of AN/TPS-3 receiver.

box described in Sec. 186. Since useful r-f gain can be obtained at 600 megacycles, two r-f amplifiers are provided. These are type 2C40 disk-seal triodes, operated as grounded-grid amplifiers. The local oscillator is an acorn tube, type 955. In the interest of low noise, a silicon crystal mixer is used in the plate circuit of the second r-f stage. The local oscillator signal is injected in the cathode circuit of the same stage.

A six-stage i-f amplifier, having a gain of about 110 db and an over-all bandwidth of about 1.25 megacycles, follows the mixer stage. Gain control is provided by manual adjustment of grid bias on the first and second stages. No automatic-frequency or gain-control circuits are used. A diode second detector and two-stage video amplifier complete the circuit. The noise figure of this receiver is about 10 db above the theoretical limit. The pulse length is 1.5 μsec and the over-all bandwidth 1.25 megacycles (relative bandwidth about $1.9/d$).

197. Typical Microwave Receivers. SCR-584 and AN/MPG-1.—A microwave receiver typical of 10-cm practice is that of the SCR-584 gunfire-control radar. The preceding r-f elements of this radar, including the t-r box, have been described in previous chapters. The block diagram of the SCR-584 receiver is shown in Fig. 378. It employs a silicon crystal in a pencil type mixer and a reflex klystron local oscillator (type 721A). The local oscillator signal is 30 megacycles higher than the carrier value, and the 30-megacycle intermediate frequency

is amplified in two i-f stages located directly adjacent to the mixer (Fig. 379). Automatic frequency control is not provided but manual adjustment of frequency is available through a potentiometer in the repeller electrode circuit of the local oscillator.

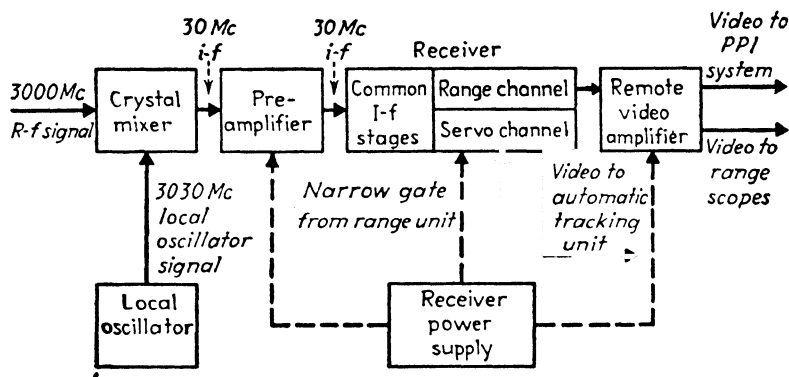



FIG. 378.—Block diagram of SCR-584 receiver.

The main group of i-f stages, second detector, and video stages are shown in Fig. 380. Single-tuned interstage coupling is used with 6AC7 pentode tubes. The shunt loading of the interstage circuits is about 800 ohms and the tube transconductance is about 9,000 micromhos resulting in a stage gain of about $0.009 \times 800 = 7.2$ per stage. The over-all bandwidth of the i-f amplifier is about 1.7 megacycles, and the pulse width is 0.8 μsec (relative bandwidth is about $1.4/d$). The over-all noise figure of the receiver is 15 db above theoretical.

Two i-f channels are provided, the two being common up to the output of the fifth i-f stage. At this point the channels separate into the range channel and the servo channel. The servo channel feeds the automatic tracking circuits and consequently must be gated against all incoming echo signals except that from the selected target. The positive narrow gate signal, generated in the timing equipment (Sec. 177), performs this function by raising the screen grid of the sixth servo i-f stage to operating potential during the gate interval. After further i-f amplification in a seventh stage, the servo signal is detected, the cathode and grid of one section of a double triode acting as a diode detector. The detected signal is passed to a triode video amplifier through a series peaking coil and thence to an output video

FIG. 379.—Simplified schematic diagram of SCR-584 preamplifier.

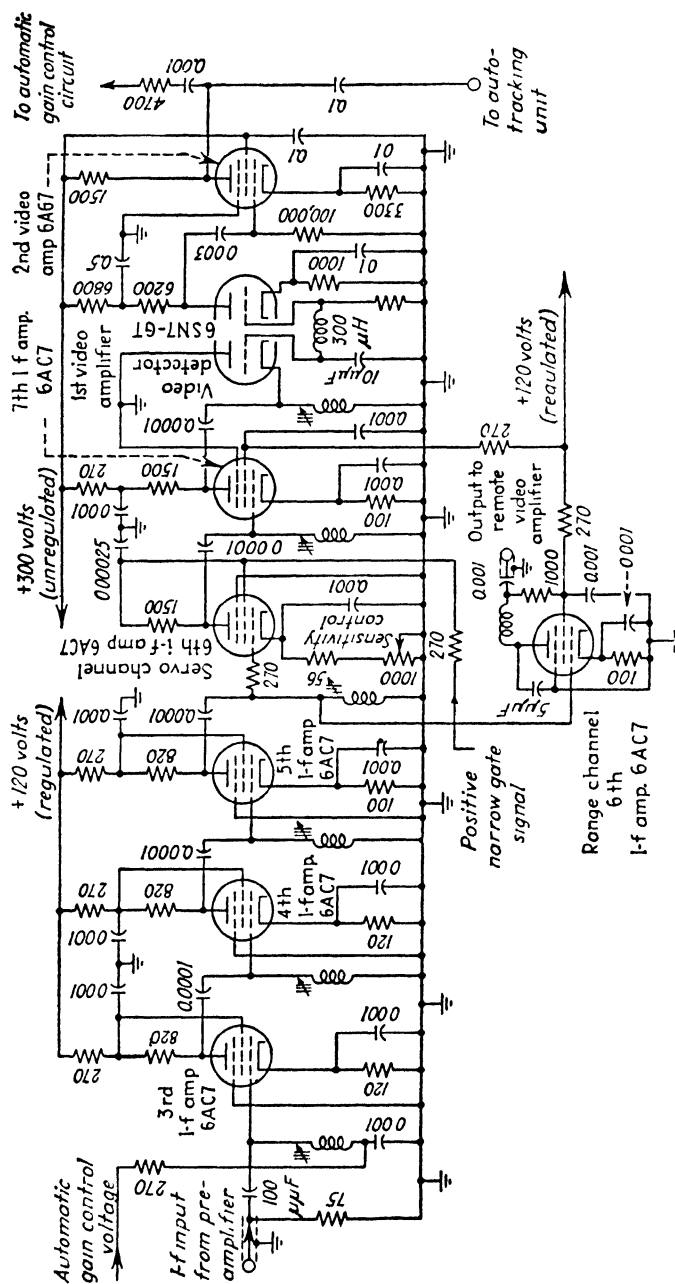



FIG. 380.—Simplified schematic of SCR-584 receiver, following preamplifier circuit.

stage employing a 6AG7 beam power tube. Since only the selected echoes appear in this circuit, it is suited to the automatic gain-control function. The video output is passed to a peak detector and d-c amplifier that acts on the second and third i-f stages, maintaining the amplitude of the selected echo approximately constant as it moves toward or away from the radar. This equalizing of amplitude is particularly important in the automatic tracking function, since the tracking circuits respond

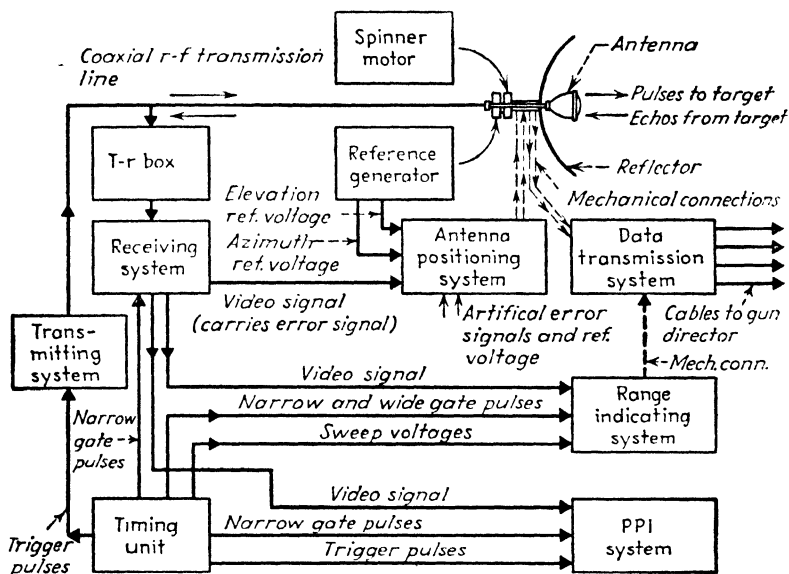



FIG. 381.—Block diagram of SCR-584 automatic tracking system.

to the amplitude of the error signal that must display amplitude variations due to angular deviations only, not due to fading or other propagation effects. Manual control of gain is provided by varying the cathode bias resistor of the sixth servo i-f stage.

The second i-f channel feeds the visual indicators (ppi and type J scopes) previously mentioned (Sec. 177). It consists of a sixth i-f stage which operates without gate control. The plate circuit of this stage serves as a detector and the video output is passed through a coaxial cable to a remote video amplifier in the indicator equipment (Sec. 211).

The final example of receiver practice, that of the AN/MPG-1

radar, is noteworthy for the number of different features it illustrates. The receiver operates at approximately 10,000 megacycles, employs stagger-tuned i-f stages with over-all bandwidth of 10.6 megacycles, and has a rapid-acting afc circuit as well as a continuous sensitivity-time-control circuit. The block diagram of the receiver is shown in Fig. 382. The t-r and anti-t-r elements have been described in Sec. 188. The echo signals from the t-r box are applied to the silicon crystal mixer (type 1N23A), where they are mixed with the output of the reflex

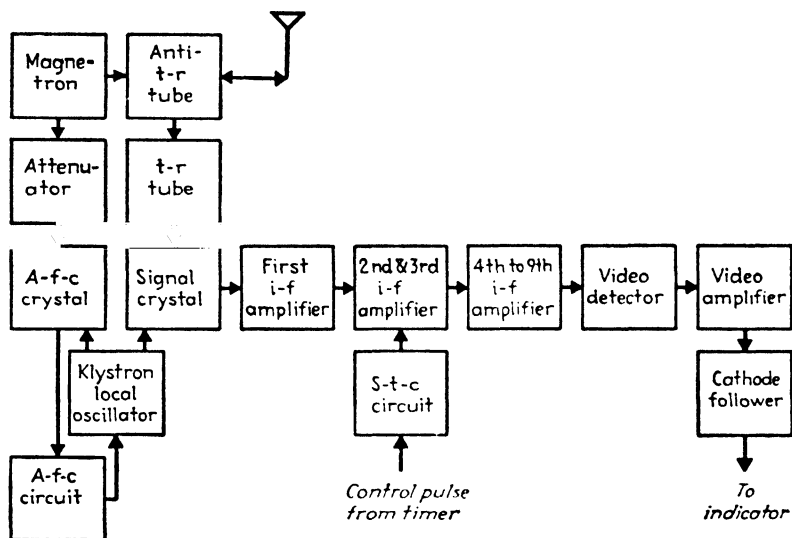



FIG. 382.—Block diagram of AN/MPG-1 receivers.

klystron local oscillator (type 723A or 723B), producing a 30-megacycle intermediate frequency.

The i-f signal is amplified in nine cascaded stages, each with single-tuned interstage coupling. Stagger tuning is used at three frequencies: 24.8, 30, and 36.3 megacycles. The nine stages are tuned in groups of three to these values as shown in Fig. 383. The net response of three stages and the over-all response of the nine stages are also shown in the figure. The over-all bandwidth at the half-power points is 10.6 megacycles, representing a relative bandwidth of about $2.7/d$ with the 0.25-μsec pulse. The noise figure of the receiver is 17 db above the theoretical limit.

The gain of the i-f amplifier is controlled continuously on a time basis by a sensitivity-time-control waveform applied to the grids of the second and third i-f stages. The control waveform has a trapezoidal shape and is generated by the circuit shown in Fig. 384. A trigger from the timer controls a sawtooth generator at the pulse rate. The sawtooth output is converted to trap-

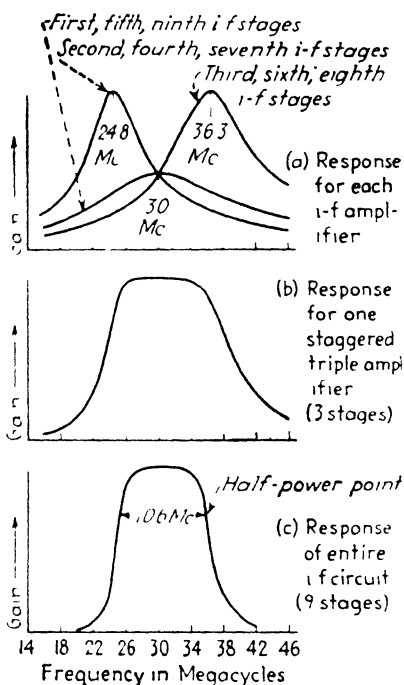



FIG. 383.—Derivation of AN/MPG-1 i-f amplifier bandpass characteristic from stagger-tuned stages.

ezoidal shape in a limiter and the peak of the limited wave is referred to zero potential by a d-c restorer diode. The control wave extends over a major portion of the pulse interval. Its amplitude is adjusted manually. The wave increases the gain of the i-f amplifier almost linearly for the first several hundred microseconds following each transmitted pulse. Full gain continues thereafter until the end of the control pulse (about 500 μsec after the transmitted pulse).

The automatic frequency control of the MPG-1 receiver is illustrated in Fig. 385. The control signal is taken from the

transmitter waveguide through a 70-db attenuator and fed to a separate silicon crystal. The use of the attenuator avoids the possibility of enemy interference with the tuning of the radar, since a strong c-w (continuous-wave) jamming signal, received through the conventional receiver channel, might take control

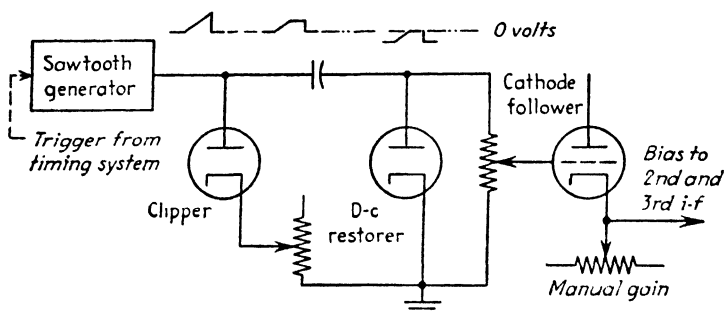



FIG. 384.—Elements of AN/MPG-1 sensitivity-time-control (s-t-c) circuit.

of the local oscillator. The attenuator effectively removes any such signal while permitting a sufficient amount of the transmitted signal (which does not pass through the t-r box in this case) to actuate the afc circuit. The auxiliary afc

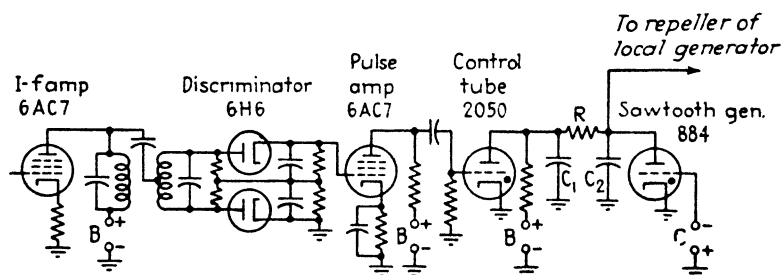



FIG. 385.—Simplified schematic of AN/MPG-1 a-f-c circuit.

crystal (type 1N23B) receives the output of the receiver klystron local oscillator and develops a 30-megacycle intermediate frequency. This is passed through a single i-f stage and thence to a discriminator and video amplifier. The output of the discriminator has a short time constant, so the output consists of pulses that are positive, zero, or negative depending on whether the intermediate frequency is below, on, or above the assigned value.

The repeller electrode of the klystron is connected with a sawtooth generator, a type 884 gas triode in a simple capacitance-discharge circuit. So long as the voltage on the anode of the 884 can rise to 100 volts or more, this circuit continues to generate a sawtooth wave of about that magnitude, which swings the klystron through a range of 100 megacycles once every 2 sec. If the receiver is well off tune this wide swing is sufficient to cause the intermediate frequency to pass through its assigned value. When it does so, the discriminator output produces positive pulses that actuate the grid of the 2050 gas tetrode control tube. This tube thereupon breaks down, shortly after the intermediate frequency passes through the correct value. The breakdown of the 2050 tube removes the voltage from the 884 sawtooth generator, which thereupon ceases to conduct. The voltage on the repeller electrode thereupon tends to remain steady, producing the correct value of the intermediate frequency. Small variations are corrected by the 2050 control tube, which becomes nonconducting when the discriminator pulses are negative. This allows the plate capacitor to charge, varying the frequency until the discriminator pulses become positive, whereupon the control tube breaks down once more. The sawtooth generator takes hold should the deviation become so large that the frequency falls outside the limits of the discriminator so that the control tube could not correct it. The remaining stages of the receiver are conventional, consisting of a diode second detector, video amplifier, and cathode-follower output stage.

CHAPTER XI

INDICATORS AND SCANNERS

We come now to the final elements of a radar system, (1) the scanner which searches for targets and determines their angular coordinates and (2) the indicators which present these coordinates and the range information supplied by the timer and receiver. These elements are peculiar to the radar art; unlike the transmitting and receiving equipment they have no direct counterparts in conventional radio engineering. The scanner derives from other target detecting and control mechanisms such as the mounts of guns, searchlights, and acoustical detectors. The indicator is similar to the picture tube of a television receiver.

198. Components and Functions of Scanners and Indicators.— The functions of the radar scanner are as follows:

1. To direct the radiated beam along a prescribed path in spherical coordinates, imparting motions in elevation and azimuth in the relationship required to perform the desired scanning pattern (circular, helical, conical, spiral, etc.).
2. To impart the azimuth and elevation motions at the rates prescribed in the scanning pattern.
3. To develop electrical or mechanical quantities functionally related to the azimuth and elevation angles, and to impart these quantities to the corresponding mechanical or electrical indicator.
4. To perform the above functions with a degree of accuracy appropriate to the intended use of the radar (cf. search function with gunfire-control function).

The components which perform these functions include motor drives (usually electrically motivated), which are geared to a mounting capable of independent motions in azimuth and elevation. This mounting supports the radiator. Connected to the shafts of the mount are potentiometers that develop electrical voltages or currents proportional to the corresponding azimuth or elevation angles. In other equipments, a direct mechanical connection is made between the shaft and a pointer visible to the operator. In still other radars, the transmission of angular

indications is through a selsyn indicator system connected to a pointer or cathode-ray indicator. In gunfire-control radars it is necessary to impart to the scanner motions derived by automatic or manual control from the target indication. This function is generally carried out by electrical means, such as the selsyn and amplidyne. When such systems are employed, means must be found to assure smooth following of the target, that is, to avoid hunting.

The functions of the indicator are

1. To provide an indication of the presence of an echo signal. This indication is usually of the visual type.[1]

2. To provide echo indications with a system of such low inertia that impulses of the order of a microsecond in width can be clearly resolved.

3. To convert the target information (range from receiver and timer, azimuth and elevation from scanner) to corresponding motions, or changes in brightness, of the visual indicator.

4. To present the target coordinates in a manner appropriate to the scanning motion employed and to the intended use of the radar. Automatic conversion from target indication to gun-control data is generally required in gunfire-control systems. This implies a design intimately related to the gun director used.

These functions of the indicator are carried out in virtually all cases by a cathode-ray indicator tube. This device enjoys its popularity for one reason only: the stream of electrons within it possesses so small an inertia that it can follow the very rapid shifts of voltage associated with pulses of 1 μsec (or less) in width. The electronic stream, modulated in intensity or deflected in position by the echo signal, is transformed to a visual indication by impinging on a fluorescent material ("phosphor"). The properties of the phosphor may be chosen to meet particular needs of the system, particularly persistence from one target observation to the next.

The target coordinates (range, elevation, and azimuth) are indicated on the phosphor screen by imparting motions to the cathode-ray beam that bear some direct relationship to the coordinates. Generally it is not convenient to display all three

[1] The only important exceptions are the Doppler effect radar and the f-m radar, both of which may give an audible indication of the presence and general direction of the target.

coordinates on a single indicator, so two or more indicators are provided. In exceptional cases all three coordinates may be shown in a single tube, at some loss in ease of interpretation.

Associated with the cathode-ray tubes are auxiliary circuits, notably the power supply, sweep circuits, and intensifier circuits. Two types of sweep circuits are used, the range sweep and the direction sweep. The range-sweep circuit imparts a motion to the cathode-ray beam that starts with the transmitted pulse and continues linearly thereafter for a time that exceeds the echo interval at the greatest range of interest. Thereafter the range sweep returns the cathode-ray beam to its starting point where the beam rests until the ensuing transmitted pulse appears and the deflection is repeated.

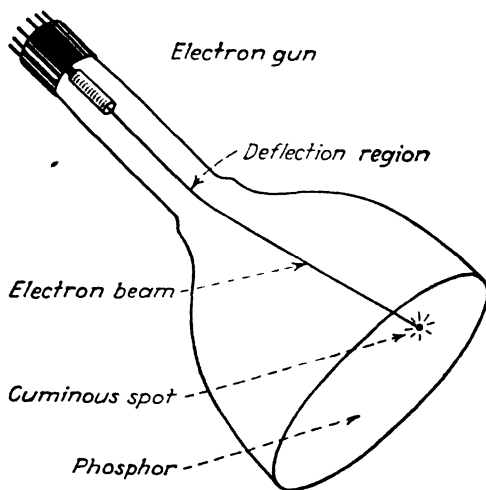



FIG. 386.—Cathode-ray indicator.

The direction-sweep circuits are connected to potentiometers or selsyns geared to the shafts of the scanner and impart to the cathode-ray beam additional motions that indicate the azimuth and elevation angles of the radiated beam relative to specified reference axes.

The intensifier circuits (also known, by their converse function, as "blanking circuits") brighten the cathode-ray beam during the active phase of the sweep and extinguish it at all other times. This action reduces the noise which would otherwise be visible on the screen, prevents confusion between forward

and retrace motions in each sweep, and generally limits the visible indication to the time during which the indicated value of range actually applies.

The power-supply circuits provide the necessary alternating and direct voltages to energize the cathode-ray tube and the auxiliary circuits. Of particular importance is the high voltage that accelerates the electrons in the cathode-ray beam. The high voltage (generally of the order of 5 kv) interposes problems of insulation and corona prevention that are particularly difficult in airborne applications. The high voltage also presents a hazard against which appropriate precautions must be taken.

199. The Cathode-ray Indicator Tube.—The cathode-ray tube (Fig. 386) is a funnel-shaped evacuated glass envelope within

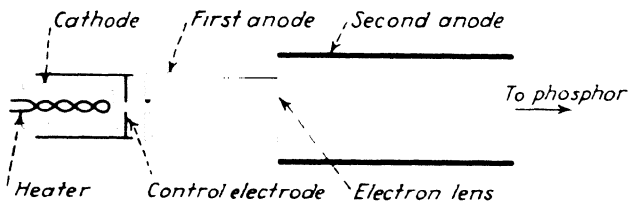



FIG. 387 Typical electrostatic electron gun.

which are the electron gun, which forms the cathode-ray beam, and the phosphor, which converts the electronic energy into visible energy. When electric deflection is used, the tube also contains the deflection system, two sets of plates symmetrically disposed about the beam. When magnetic deflection is used, the deflection system is external to the tube, in the form of one or more sets of magnetic deflection coils known as a "deflection yoke." The electron gun may also have an external magnet coil for focusing the beam. In all cases the electron gun has an electrode for controlling the beam current density and, consequently, the brightness of the trace on the phosphor.

Two types of electron guns are commonly employed. The electrostatic-focus gun is illustrated in Fig. 387. It consists of a flat or cup-shaped cathode coated with oxides of barium and strontium and heated indirectly by a tungsten wire. Adjacent to the cathode surface is the control electrode (often referred to as the "grid"), through an aperture in which the electron beam is initially formed. The potential on this electrode, relative to the cathode, has a marked effect on the current in the beam,

as well as a secondary effect on the cross-sectional area of the beam.

Next adjacent to the control electrode is an accelerating electrode, the first anode, also containing an aperture. This anode serves to accelerate the electrons passing through the control electrode, and constitutes the first element of an electron lens. The second element of the lens is the second anode, which usually takes the form of a conducting coating on the inside surface of the tube envelope. The outer edge of the first anode lies in the same plane as the inner edge of the second anode. The electron lens forms an image of the cross section of the electron beam at its minimum area, the so-called "crossover," which is generally situated slightly in front of the cathode surface. By adjusting the potential on the first anode, relative to the cathode,

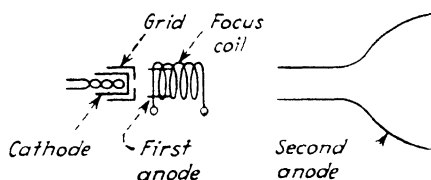



FIG. 388 -Magnetostatic-focus electron gun

the focal length of the electron lens can be adjusted until the spot of light produced on the phosphor has its minimum size. The potential on the control electrode (grid) controls the brightness of the spot.

A modern version of the electrostatic electron gun is the so-called "zero first-anode current" type. In this gun the first anode does not have defining apertures and hence does not intercept electrons in forming the beam. Since the first anode does not consume current, beam current variations caused by the control electrode do not affect the focus of the beam, and the anode power supply need not have good regulation.

The magnetostatic electron gun (Fig. 388) consists of a coil of wire wound axially around the neck of the cathode-ray tube. The focusing action results from the fact that electrons originating at the crossover, and emerging from it at an angle to the tube axis, are caused by the axial magnetic field to travel through a spiral which completes one turn in the length of the tube, regardless of the angle at which the electrons emerge. Hence, all

electrons leaving the crossover may be brought to focus at the phosphor screen.

The magnetic gun has a cathode, control electrode, and second anode like the electrostatic gun, and the brightness of the beam is controlled in the same manner. The magnetic focusing field must have uniform distribution over the region within which the beam travels, to avoid defocusing the beam as it is deflected. The field must also be constant with respect to time.

200. Phosphors.[1]—The material that transforms the energy of the electron beam into light is known as the "phosphor." Phos-

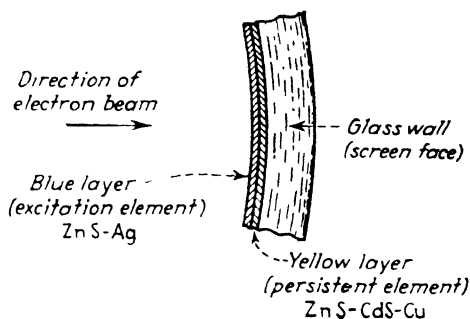



FIG. 389.—Cascade (long-persistence) phosphor.

phors have a variety of characteristics, depending on their chemical and physical composition. The two types most widely used in radar indicators are the single-layer (short-persistence) phosphor and the double-layer cascade (long-persistence) phosphor. The single-layer phosphor produces green light of rapid decay and is used primarily in type A indicators for indicating pulse waveforms and in tubes intended for use with high-speed scanners. The cascade phosphor is used where persistence of the image is required, as in types B, C, ppi, and similar indicators.

The single-layer green phosphor is composed of zinc silicate (willemite) with a manganese activator. It fluoresces with a brilliant green light, it has high electrical and visual efficiency, and the light output decays to 1 per cent of its initial value in about 0.05 second.

The cascade phosphor (Fig. 389) consists of two layers of radically different properties. The layer on which the electrons

[1] LEVERENZ, H. W., Luminescence and Tenebrescence as Applied in Radar, *RCA Rev.*, **7**, 199 (June, 1946).

first impinge (that next to the electron gun) is made of zinc sulphide, activated with silver. It fluoresces with a brilliant blue light, which decays a few milliseconds after removal of the excitation. The second layer, next to the glass envelope of the tube, is composed of zinc sulphide and cadmium sulphide, with a copper activator. This layer is excited by the blue flash from the first layer and glows with a yellow color that persists for several seconds. The light from the second layer remains visible for several minutes when viewed in total darkness. The second layer is not excited to any appreciable degree by the electron beam since the electron energy is absorbed by the first layer. Electronic excitation of the second layer is, in fact, undesirable since the light produced by electron impact has

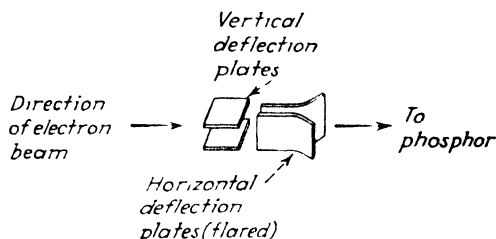



FIG. 390.—Flared-plate electric deflection system.

much shorter persistence than that induced by the blue light of the first layer.

To avoid eyestrain it is desirable to remove the brilliant, rapidly decaying blue light produced by the first layer. This is accomplished in part by the second layer, which has a yellow color and hence absorbs the blue light. As a further precaution, radar indicators are usually fitted with an amber-colored optical filter that removes the remaining blue light without intercepting any substantial amount of the yellow light.

201. Deflection Systems.—Two methods are employed to deflect the electron beam over the surface of the phosphor: electric deflection and magnetic deflection. In the electrically deflected system (Fig. 390) the cathode-ray tube is fitted with two sets of deflection plates arranged within the neck of the tube on either side of the beam. The deflecting plates exert force on the electrons at right angles to their motion, along two mutually perpendicular directions, as shown in the figure. The deflection experienced by the beam is proportional to the voltage across

the deflecting plates and is inversely proportional to the separation of the plates and to the velocity of the electrons in the beam. In general, the voltage required to produce a given deflection increases directly with the second anode voltage.

To assure the highest possible deflection sensitivity, that is, the maximum deflection per volt applied to the deflection plates, it is customary to flare the plates outward in the direction of the beam, as shown in Fig. 390. This permits the plates to be placed close together, and hence to exert a maximum electric field on the beam, without causing them to intercept the beam at maximum deflection. Electric deflection is used in type A indicators

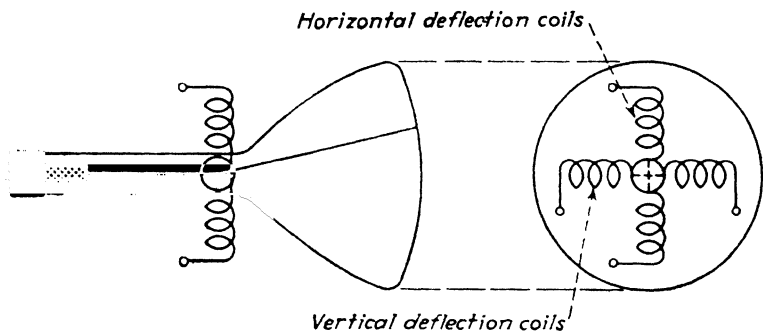



FIG. 391.—Coil system for magnetic deflection.

and to a limited extent in type B and ppi systems. The latter tubes must usually operate at high second-anode voltage to secure adequate brilliance from the cascade phosphor. The high second-anode voltage requires correspondingly high deflection voltages, which are difficult to generate.

As a further aid to deflection sensitivity, the electrons may be accelerated after deflection by a so-called "intensifier electrode." This takes the form of a ring on the inside of the tube next to the phosphor. A brighter image may be obtained in this way without appreciably reducing the deflection sensitivity.

The magnetic deflection system (Fig. 391) is preferred because the required deflection can be obtained, at high accelerating voltages, by high current rather than by high voltage. The high current requirement implies deflection amplifier tubes of correspondingly high current drain, but the problems of high voltage insulation and corona prevention are minimized.

The magnetic system consists of one or more pairs of coils wound on an axis at right angles to the electron beam. The coils are formed to fit around the neck of the tube, and the turns are distributed so as to produce a uniform field across the neck of the tube. When current is applied to the coil, the beam is deflected in a circular path around the magnetic lines of force. The beam thus travels in an arc of a circle, and it emerges from the deflecting field traveling at an angle to the axis of the tube, eventually striking the phosphor at a point off the center of the screen. Two sets of coils, arranged with mutually perpendicular fields, can cause the spot to assume any position on the screen.

The extent of the magnetic deflection increases proportionately as the current in the coils increases, and it is inversely proportional to the square root of the accelerating voltage. Thus

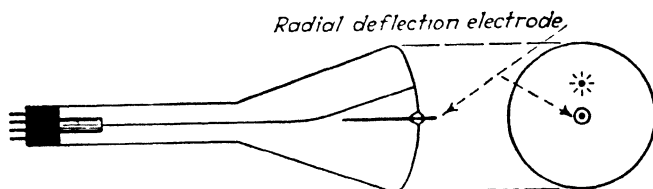



FIG 392.—Radial defection by electric system.

large deflections are more readily attained, with high accelerating voltages, by the magnetic method than by the electric method of deflection.

The magnetic deflection system suffers from the formation of a blackened spot in the center of the phosphor. This spot is caused by the gradual desensitizing of the phosphor under the effects of a steady ionic bombardment. Negative ions (massive negatively charged atoms) emitted from the cathode are accelerated through the electron gun and impinge on the phosphor. In electric deflection systems, the ions and the electrons are deflected to the same extent, so no local desensitizing occurs. But in magnetic systems the deflection is inversely proportional to the mass of the cathode-ray particles, and the massive negative ions, suffering correspondingly little deflection, bombard the center of the screen. A number of systems have been devised for obviating this effect. One method is the ion trap, in which the electron gun is bent off the axis of the tube and a magnetic field is applied within the electron gun to deflect the

electrons (but not the ions) back to the axis. Another system is the use of a metal-backed phosphor, that is, a thin layer of metal atoms deposited on the phosphor. The electrons penetrate the metal layer and excite the phosphor, but the more massive ions cannot do so and are conducted away by the metallic coating.

A specialized form of electric deflection system is the radial deflection electrode shown in Fig. 392. This consists of a wire extending part way down the axis of the tube through a seal at the center of the screen. A negative deflecting voltage applied between this electrode and the cathode (or some other common element of the electron gun) causes the beam to be deflected away from the center of the tube, as is required in the

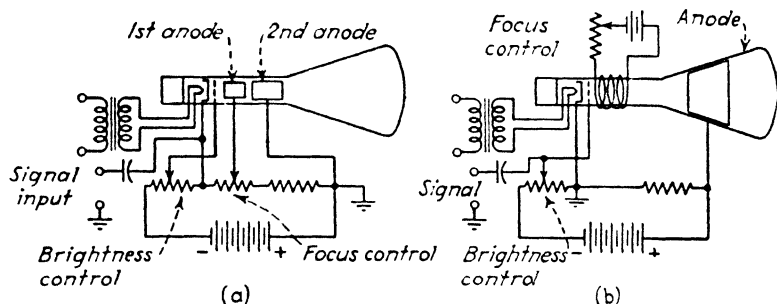



FIG. 393.—Auxiliary circuits for cathode-ray indicators: (a) electrostatic gun; (b) magnetostatic gun.

type J indicator. Conventional electric deflection plates are also included in this type of tube to provide circular deflection of the spot about the center of the tube.

202. Auxiliary Circuits for Cathode-ray Tubes.—The power supply and electron gun circuits of typical magnetic and electric systems are shown in Fig. 393. The essential controls are the focus control, manual and automatic brightness controls, and astigmatism control. The focus is controlled in the electrostatic gun by varying the potential on the first anode, relative to the cathode (or other electrode of fixed potential). In the magnetic gun, the focus is found by adjusting the current in the focus coil. A resistive voltage divider across the high-voltage power supply provides the first-anode and second-anode voltages, as well as the negative bias on the control grid electrode (or positive bias for the cathode).

As we shall see in the following discussion of deflection systems, it is important in electric deflection that the average potential of the deflection plates bear a close relationship to the second-anode potential. Otherwise, the net electric force acting on the beam varies with the position of the beam and defocusing occurs. Thus it is highly desirable to employ push-pull deflection, with the average deflection potential referred to the second anode level. Certain aberrations of focus can be corrected by adjusting this potential relationship. If the electron gun and deflection system display astigmatism, a potentiometer, known as an "astigmatism control," may be inserted to vary the average deflection voltage relative to the second anode.

Control of brightness is effected by changing the potential between the cathode and the control grid, the spot becoming brighter as the control grid becomes more positive. The choice of circuit depends on the polarity of the signal as derived from the receiver. If the signal polarity is negative, it is applied to the cathode, and the control grid, by-passed to ground, serves as the brightness-control electrode. If the pulse signal is positive, it is applied to the control grid, and the cathode, properly by-passed, serves as the brightness control. Any electrode may be connected to d-c ground, depending on circuit requirements. Generally, the cathode is grounded, but in some cases the second anode is grounded as a safety precaution. In the latter case the high accelerating potential is applied to the cathode in negative polarity.

A difficulty in the brightness-control circuit is the tendency of the spot to become defocused when the amplitude of the signal exceeds a certain limit. This effect, known as "blooming," may destroy the resolution of the indicator when strong echoes are received. It can be avoided by a brightness-limiting circuit, a limiter that prevents the echo signal from attaining a value above the blooming level. A diode limiter may be attached directly to the control electrode or cathode (depending on which serves as the signal electrode). Alternatively the video amplifier stage feeding the cathode-ray tube may be operated as a grid-voltage limited stage so that high-amplitude signals cut off the amplifier. The positive peak value of signal appearing at the plate of the amplifier cannot exceed the corresponding limit, and blooming is thereby prevented.

Another aspect of the signal circuit requiring special treatment is the d-c component of the pulse signal. When, as is usual, the second detector and video amplifier stages are capacitively coupled, the d-c component of the signal is removed and the *average* value of the pulse waveform is superimposed on the bias level in each stage. Consequently, when a strong echo is received, the peak values of other signals in the field of view are depressed and may become lost to view unless the manual brightness control is adjusted to compensate for the change. Such manual adjustment is not practical during scanning; therefore it is usual to restore the d-c component of the signal prior to its application to the indicator tube. The d-c component is restored by passing the echo pulses through a diode that

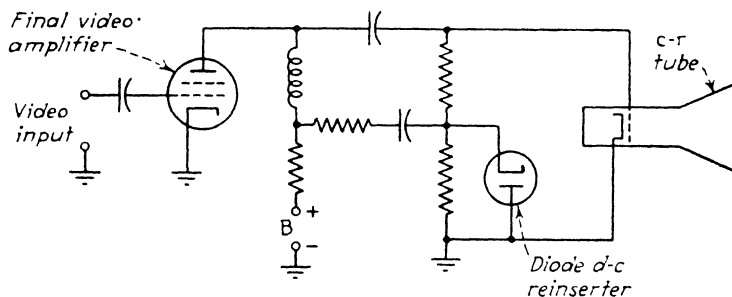



FIG. 394.—Signal circuit of c-r indicator, showing diode for d-c reinsertion (automatic brightness control).

charges a capacitance to the peak value of the received voltage wave. The direct voltage across the capacitance is made part of the brightness-control bias of the c-r tube. The *RC* time constant of the circuit is made short compared with the scanning interval, so that bias adjusts itself in accordance with the peak value of the echo signals. The minimum brightness of the indicator screen is thereby kept constant irrespective of the changes in echo signal level as the scanning proceeds. A typical automatic brightness control circuit is shown in Fig. 394.

203. Deflection Patterns for Various Types of Indication.—As we recall from Chap. I (cf. Table III) a variety of deflection patterns are employed in the different types of radar indicators. The basic types are A, B, C, ppi, and J. In the type A sweep, Fig. 395, the deflection pattern is rectangular, range indicated by horizontal deflection of the beam and signal intensity by vertical

deflection. The signal does not modulate the beam intensity, but manual brightness control and intensifier pulses are provided. The primary deflection requirement of the type A indicator is a linear horizontal sweep, which starts with the transmitted pulse and lasts for the maximum desired echo interval. A high degree of linearity is required when the distance units are marked on a

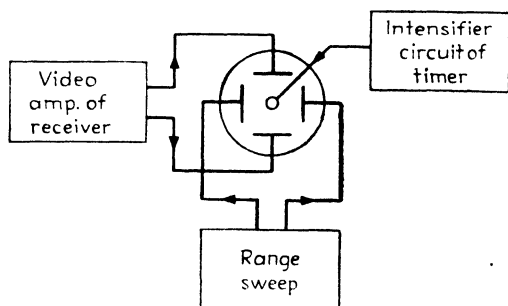



FIG. 395.—Deflection system for type A indicator.

fixed scale attached to the indicator. If range markers are used, some nonlinearity may be tolerated.

In the type B indicator (Fig. 396) the range sweep deflects the beam vertically. The requirements are the same as in the range sweep of the type A indicator, namely, a high degree of linearity, starting time coincident with the transmitted pulse,

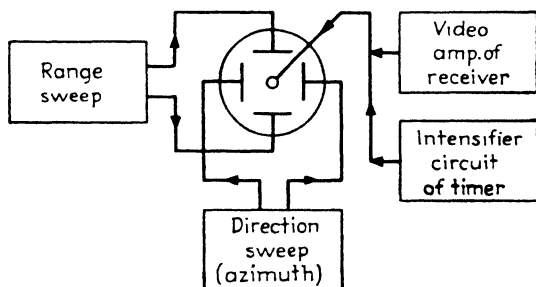



FIG. 396.—Deflection system for type B indicator.

and duration of the sweep sufficient to cover the maximum echo interval on the range scale then in use.

The horizontal sweep of the type B scan is derived from the angular position in azimuth of the radiated beam. The sweep current (or voltage), obtained from a potentiometer mounted on the azimuth shaft of the scanner, is directly propor-

tional to the azimuth angle of the radiator, relative to a fixed reference (for example, true north). A strictly linear relationship between azimuth angle and horizontal deflection must be maintained if, as is usual, the angle scale on the indicator has fixed and equal intervals. Since such linearity is not difficult to obtain (by introducing the required taper in the potentiometer winding), angle-marking pulses are not required, although they are provided in some equipments.

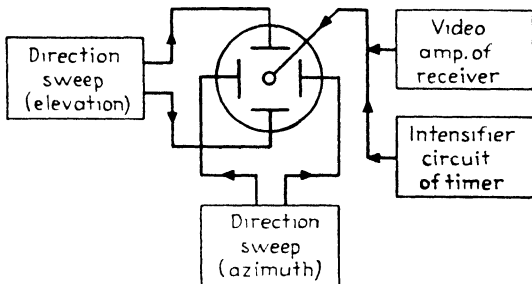



FIG. 397. Deflection system for type C indicator.

In the type C indicator (Fig. 397) two angle deflections are used, elevation vertically and azimuth horizontally. The deflection currents are derived from two potentiometers affixed to the corresponding shafts of the scanner. Each potentiometer has the proper taper to provide a linear relationship between scanner angle and the position of the indicator beam. As we have seen in Sec. 17, the type C indicator requires a gating pulse that intensifies the beam during the reception of the desired echo and blanks it out at other times.

In the type B and C indicators magnetic deflection is customarily used, not only because it permits adequate deflection at high accelerating voltage (and hence at adequate brightness), but also because the potentiometers that develop the angle-deflection currents then may be operated at low voltage. This is a decided advantage since the insulation requirements are reduced, and the entire angle-deflection system may operate at low impedance, through coaxial cables, without the need of amplification.

The ppi sweep (Fig. 398) has a range sweep not radically different from those employed in the type A and type B indicators. The same range marking and intensifier pulses are

used. The deflection pattern is complicated by the fact that the direction of the radial deflection must correspond to the azimuth direction of the scanner. Magnetic deflection is customarily used, and the rotating deflection may be obtained in one of two ways. In the first, the magnetic scanning yoke (a single pair of coils fed with the range-sweep deflection current) is rotated mechanically about the neck of the indicator tube by a mechanical or electromechanical connection with the azimuth shaft of the scanner. In the second method, the scanning yoke is fixed, and two sets of coils are provided at right angles to each other. The amplitudes of the range-sweep currents in the two sets of coils are modulated with sinusoidal envelopes in quadrature.

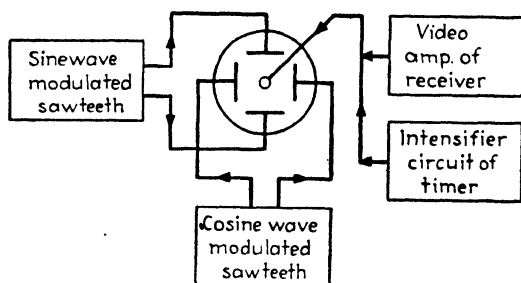



FIG. 398.—Deflection system for plan position indicator.

The vector sum of the two deflections causes the beam to move along a radius whose direction corresponds to the azimuth angle of the radiated beam. A number of auxiliary requirements, which must be satisfied in this fixed-coil ppi sweep, are discussed in detail in Sec. 205.

The type J sweep (Fig. 399) is used primarily for precise range measurements. Since the required deflection is not great, and since persistence is not required, electric deflection is customarily used. A circular Lissajous figure is provided by sinusoidal voltages applied to the deflection plates in quadrature. On this circular sweep is the signal deflection superimposed, in the radial direction, by application of the echo pulses to the radial deflection electrode. A high degree of linearity in the circular sweep can be obtained by maintaining the purity of the sinusoidal sweeps (eliminating harmonics). Two type J sweeps are sometimes used in conjunction, one to show the full range interval of interest, and the other (vernier scope) to show

in greater detail the segment of range within which the target lies. Typical type J scope circuits are described in Sec. 211.

204. Linear Sweeps for Range Measurement.—Linear sawtooth waves of voltage or current are required for range measurement, voltage waves for electrically deflected tubes, and current waves for magnetically deflected tubes. The sawtooth wave of voltage is generated across a capacitor, amplified and applied without change in shape to the deflecting plates. The sawtooth wave of current is generated in an inductance or tuned circuit, amplified, and applied to the deflecting coil.

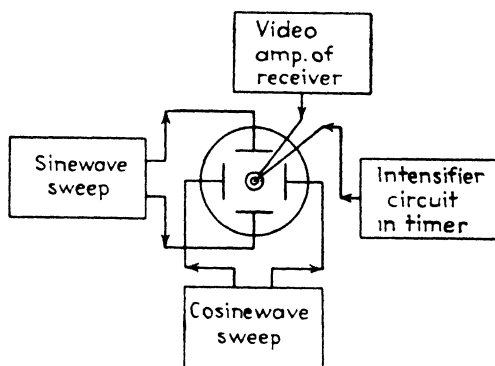



FIG. 399. Deflection system for type J indicator.

Either of two methods may be used to generate sawtooth voltage waves: (1) charging (or discharging) a capacitor through a resistance, employing only the early portion of the exponential rise (or fall) of voltage, or (2) discharging a capacitor through a constant-current tube, and using the full extent of the linear increase of voltage. These circuits have been mentioned briefly in Sec. 131, Chap. VI.

The *RC* exponential circuit (Fig. 400) charges the capacitor along a curve given by

$$v = V_b(1 - \epsilon^{-t/RC}) \qquad (497)$$

where C is the capacitance charged through the resistance R, V_b is the charging voltage, and v is the voltage across the capacitor at time t, measured from the beginning of the charge. To assure reasonable linearity (voltage amplitude within 5 per cent of the linear value), it is necessary to restrict the duration of the charging period to not more than one-fifth the RC time constant.

Substituting $t = RC/5$, we obtain $v = 0.181V_b$, that is, the maximum usable voltage across the capacitor is only 18 per cent of the charging voltage. Thus if it is desired to produce a peak value of the sawtooth of, say, 200 volts (capable of producing a 2-in. deflection at a deflection sensitivity of 0.01 in. per volt), the charging voltage must be about 1,100 volts, and this high value is not readily accommodated in conventional circuits. To avoid this difficulty the sawtooth wave is generated at low voltage and then passed through a push-pull amplifier stage which, operating in linear fashion, makes full use of the available

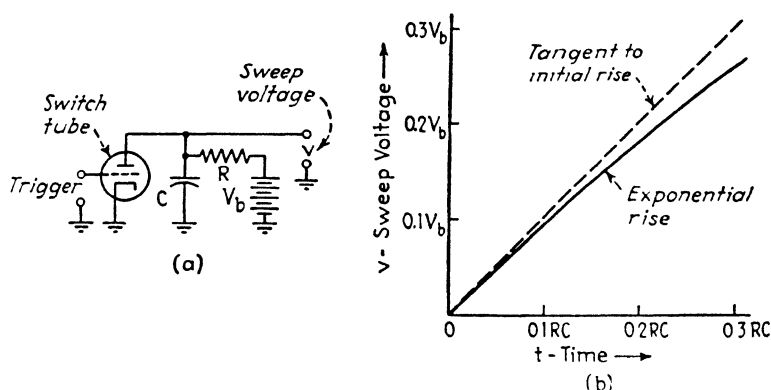



FIG. 400.—Linear deflection using initial portion of exponential charge or discharge (a) circuit, (b) resulting deflection waveform

B-supply voltage. In this manner a 200-volt peak sawtooth can be obtained from a 300-volt supply.

When it is desired to secure a closer approach to linearity than the 5 per cent tolerance mentioned above, the constant-current circuit (Fig. 401) is used. Here the capacitor is charged from the voltage source through the switch tube, which acts under the control of a rectangular wave produced by the timer equipment. Since the switch tube has low internal resistance, the charge is assumed quickly, during the quiescent period preceding the active sweep. The sweep is started when the switch tube is cut off by the negative edge of the timer pulse. At this instant, the capacitor starts to discharge through the associated constant-current pentode tube. The pentode has a flat internal impedance curve, that is, the current through it remains constant over a wide range of voltage. However, this

relationship ceases when the plate voltage falls below a critical value, and it is important that the sweep be completed before the capacitor is completely discharged. This can be arranged by proper choice of the capacitance and pentode current values.

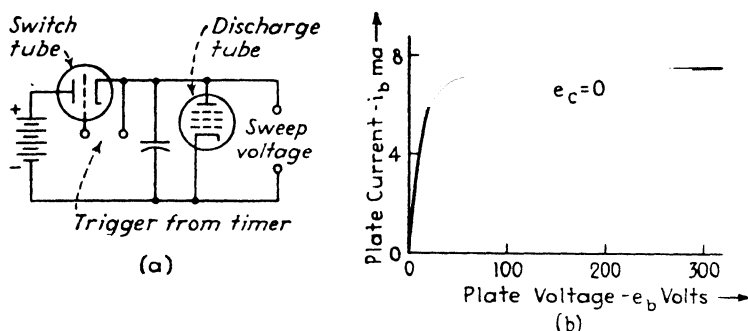



FIG. 401.—Constant-current discharge tube for improving linearity of deflection sweep (a), and constant current characteristic of pentode (b).

Up to 90 per cent of the available charging voltage can be devoted to the linear portion of the sweep in this circuit, and linearity within 1 per cent is readily achieved.

The sawtooth voltage thus generated is applied to the deflection plates through a paraphase (push-pull) amplifier (Fig. 402)

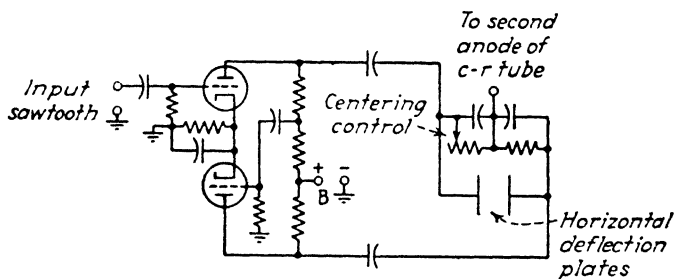



FIG. 402.—Paraphase amplifier for symmetrical application of deflection voltages.

that introduces gain if needed, and permits the average deflection voltage to be tied to the second anode value. Direct-current restoring diodes are often used directly at the deflecting plates to assure that each sawtooth wave starts from zero, rather than from the average value of the waveform. A bias value may be introduced at this point for centering the beam.

When sawtooth current waves are required, it is possible to produce an exponentially rising current through the deflection

coils by sudden application of current through a switch tube. The circuit (Fig. 403) consists of the magnetic deflection coil and an auxiliary inductance of larger value, in series with the plate of the switch tube. When the switch tube is driven to conduction, by a rectangular wave from the timer, the current in the two inductances increases exponentially, in accordance with the time constant L/R, where R is the total series resistance and L is the total series inductance. As in the case of the RC voltage generator, the duration of the sweep must be limited to a small fraction of the L/R time constant. Consequently a large

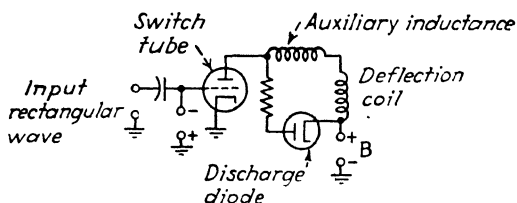



FIG. 403.—Circuit for producing sawtooth waves of current for magnetic deflection.

value of L must be used, and a large value of voltage must be available to force a large value of current through it. Consequently, low-level generation of the sawtooth, followed by linear current amplification to the required level, is often preferred. The distributed capacitance across the inductance gives rise to oscillations when the sawtooth wave is cut off, and these must be damped out by a damping diode.

A somewhat similar system makes use of a tuned circuit, of which the deflection coil is the inductive element. A capacitance placed across the coil introduces resonance at a frequency about 10 times the desired sweep frequency. When the switch tube is cut off under the control of the timer, the tuned circuit is shocked into damped oscillation. The first half cycle of oscillation may be sufficiently linear to serve as the sweep current. The circuit is complicated by the need of isolation from the B supply, so that centering current may be applied to the deflection coil. Also, the oscillations following the initial half cycle must be eliminated. This is usually done by applying blanking pulses to the control electrode or cathode.

The sawtooth current generator enjoying widest application is the trapezoidal generator shown in Fig. 404. The trapezoidal

wave of voltage can be shown to produce a sawtooth wave of current by the following argument: the deflection coil consists of inductance L and resistance R in series. To cause a linear rise of current in this combination two voltages are required simultaneously, a sawtooth voltage to produce a sawtooth current in the resistive component, and a rectangular voltage to produce a sawtooth current in the inductive component. The sum of the sawtooth and rectangular voltages is a trapezoidal wave (Fig. 404a). The trapezoid may be generated at low level and amplified in a linear high current stage that applies the wave, without change in form, to the deflection coil.

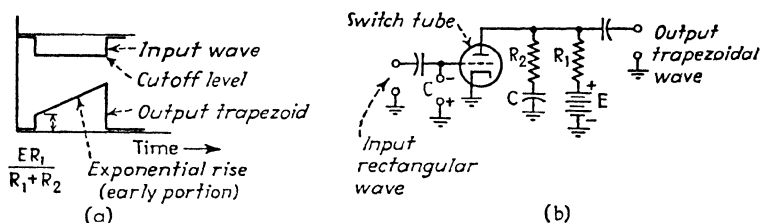



FIG. 404.—(a) Trapezoidal wave and (b) circuit for generating such waves.

The trapezoidal wave may be formed by the RC circuit shown in Fig. 404b. The switch tube short-circuits the capacitance and the resistance R, as well as the output terminals, so no voltage appears initially (except the voltage drop across the switch tube, which is small). When the switch tube is cut off by the timer pulse, the battery voltage is applied across the capacitor and the two resistors in series. The initial capacitor voltage is zero, so the battery voltage divides in the ratio R_1/R_2, and the drop across R_2 suddenly appears across the output terminals. This is the leading edge of the trapezoid. Thereafter the capacitor charges and the current through R_2 increases exponentially, according to the time constant $C(R_1 + R_2)$. This time constant is made long enough so that only the initial part of the rise occurs during the required sweep time. Hence the rise is approximately linear and the upper portion of the trapezoid wave is formed. At the conclusion of the sweep, the switch tube conducts, short-circuiting the output and forming the trailing edge of the wave. As in the case of the RC voltage sweep circuit, it is advisable to form the trapezoid at low level and to follow with amplification in a high-current stage before application to the deflection coil.

All the sawtooth current generators described above suffer from a tendency to delay the start of the sweep. This lag occurs as the stray and distributed capacitance in shunt with the scanning coil becomes charged when the driving voltage is first applied. Since the voltage across this capacitance takes time to build up, the applied voltage does not have an immediate effect on the inductance, and the sweep current is correspondingly delayed. The remedy is to limit the shunt capacitance to the smallest possible value.

The delay in the start of the magnetic sweep may be disregarded in master-oscillator timing systems by the simple expedient of starting the magnetic sweep circuit slightly in advance of the transmitter trigger, by the methods outlined in

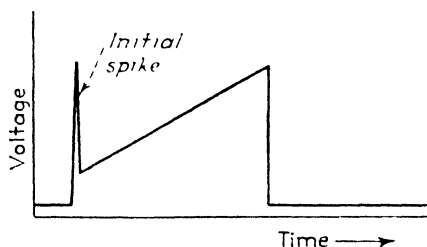



FIG. 405.—Spike on trapezoidal wave for overcoming effects of shunt capacitance at start of magnetic sweep.

Chap. VIII. In self-synchronous radar systems, however, the magnetic sweep must start slightly later than the transmitted pulse, and any further delay in the start of the sweep may preclude accurate indication of near-by targets. In such cases, it is possible to charge the shunt capacitance by a "spike" of voltage superimposed on the trapezoidal wave previously described, as shown in Fig. 405. The sharp peak of voltage is obtained from the driver circuit by passing the leading edge of a positive rectangular wave through a differentiating circuit and feeding the resulting narrow peak to the trapezoidal circuit.

205. Plan-position Indicator Sweep Systems.—The simplest method of generating the sweep for plan position indication (ppi) is the rotating yoke system (Fig. 406). The yoke consists of a single pair of coils, which deflect the beam along a line from the center of the screen outward. The d-c component of the waveform is adjusted so that the sweep starts directly at the center of the tube. The yoke is mounted on bearings and is

driven, by gears, about the axis of the tube, in step with the azimuth rotation of the scanner. Slip rings feed trapezoidal voltage waves to the deflection coil, forcing sawtooth waves of current through it. The amplitude of the sweep is sufficient to drive the beam to the edge of the screen, and the duration of the sweep corresponds to the maximum range required. The peak amplitude of the sweep remains constant.

The yoke is driven by a motor that is synchronized in some manner with the azimuth scanning motion. The most usual

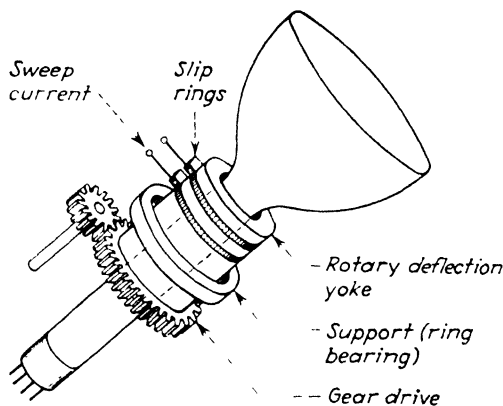



FIG. 406.—Rotating yoke method of generating ppi deflection

method is to employ a primary drive motor on the scanner, and to mount on the scanner shaft an auxiliary generator (selsyn transmitter) which develops voltages indicative of the azimuth direction. These voltages are transmitted to a similar motor (selsyn receiver) of adequate power to drive the deflection yoke. Where a high degree of accuracy is required in the angular indication, it is necessary to employ gearing between shaft and selsyn transmitter and similar gearing between selsyn receiver and yoke. The gearing permits the selsyn shafts to rotate a large number of times (say 12) for one rotation of the scanner. This arrangement permits the yoke to take up a number of synchronous positions, only one of which corresponds with the scanner direction azimuth. If 12:1 gearing is used, for example, synchronism may be established every 30 deg (= 360 deg/12) in the rotation of the yoke. To avoid this ambiguity, auxiliary contact switches may be placed adjacent to the scanner shaft and the yoke

bearing at corresponding positions, actuated by cams on each member. When one of these switches is closed, while the other remains open, the selsyn circuit is deenergized momentarily and the primary drive rotates the scanner until both contact switches are closed simultaneously, indicating that the directions indicated by the deflection yoke and scanner azimuth are the same. The system then locks in this relationship until the system is shut down or the power is otherwise interrupted.

The disadvantages of the rotating yoke method are its mechanical complexity and the difficulty of assuring rigid mounting

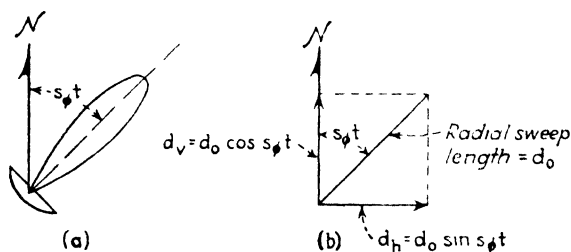



FIG. 407.—(*a*) Quantities specifying steady circular scanning and (*b*) corresponding components of ppi deflection

of the yoke, free of play in the bearings and gear drive. While the rotating yoke is used to some extent, the preferred method of ppi sweep, used in precise ground- and ship-based equipment, is the fixed yoke system described below.

The fixed-yoke ppi system employs two sets of deflection coils which deflect the beam in mutually perpendicular directions. The coils are fixed in position at the neck of the tube. Both coils are energized simultaneously, and the net deflection force is the vector sum of the two forces. By varying the relative amplitudes of the deflection currents, it is possible to deflect the beam radially from the center of the screen along any desired direction. To create the rotating sweep of the ppi, then, it is necessary to change the relative amplitudes of the two component deflections from one sweep to the next, in continuous fashion.

To investigate the required mode of variation of the two component sweeps, suppose that the scanner rotates in azimuth at a rate of s_ϕ radians per minute, starting at north as the reference. After t minutes, the beam has rotated $s_\phi t$ radians from the reference. The ppi sweep must take up a corresponding position as shown in Fig. 407. The horizontal and vertical deflections

d_h and d_v required to produce a sweep along this direction are

$$d_h = d_0 \sin s_\phi t \tag{498}$$

and

$$d_v = d_0 \cos s_\phi t \tag{499}$$

where d_0 is the peak value of the deflection. The magnitude of the resultant of these two deflections is

$$d = \sqrt{(d_0 \sin s_\phi t)^2 + (d_0 \cos s_\phi t)^2} = d_0 \tag{500}$$

Hence at all values of t, throughout the rotation of the sweep, the magnitude of the resultant deflection is a constant and equal

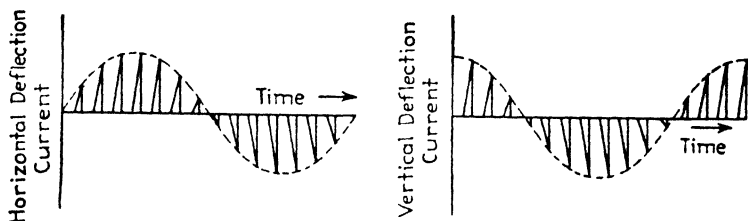



FIG. 408.—Sine-wave and cosine-wave modulation of sawtooth waves for fixed-yoke ppi deflection.

to d_0. The angle between the sweep and the north reference, $s_\phi t$, continually increases with time, that is, the sweep rotates continuously in synchronism with the scanner.

Equations (498) and (499) represent sinusoidal and cosinusoidal modulations of the sawtooth currents in each set of coils, as shown in Fig. 408. In effect, we must multiply the basic sawtooth by sine and cosine functions of the instantaneous azimuth angle $\phi = s_\phi t$. Two methods of introducing these modulations have been used in radar applications, the potentiometer method and the rotary transformer method.

FIG. 409.—Resistance card and pivoted brushes for generating sinusoidal and cosinusoidal waves.

The potentiometer modulator, shown in Fig. 409, is a flat card on which is wound a number of turns of resistance wire,

connected to a direct voltage. Contact is made with the winding by four brushes mounted on a frame that is pivoted at the center of the card. The card rotates in synchronism with the azimuth motion of the scanner. As the card rotates, the voltage picked up by each pair of brushes varies from zero (when the pair lies along a vertical line as shown) to a maximum (when the pair lies on a horizontal line). At intermediate positions, the voltage amplitude varies sinusoidally. Since the two sets of brushes are displaced 90 deg, the voltages appearing across them are sinusoids in quadrature.

The modulating potentiometer is inserted at the input of the trapezoid generating circuit (Fig. 404). Four such circuits are used, one connected to each of the four brushes, which feed

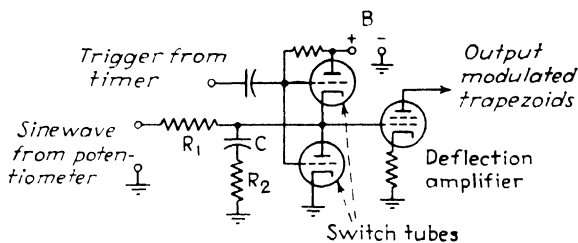



FIG. 410.—Circuit for generating sinusoidally modulated sawtooth waves from potentiometer shown in Fig. 409.

sinusoidal voltages to the trapezoid circuits as shown in Fig. 410. The start and stop of each trapezoid is controlled by switch tubes driven by the timer. Since the potentiometer reverses the polarity of its output voltage twice during each rotation, two switch tubes, connected in opposite polarity, are required to permit discharging the capacitor in either case.

The modulating potentiometer is a simple device, but is subject to mechanical wear and must be replaced from time to time. Also, the imperfect contact between windings and brushes introduces noise that must be filtered to prevent variations in the sweep amplitude.

The second fixed-yoke system, the rotating transformer, is not subject to these shortcomings, and it is widely used. In this system, shown in Fig. 411, the trapezoidal wave is generated in a low-level circuit and fed to the primary of a transformer. The primary is mounted on bearings and is rotated in synchronism with the scanner motion in azimuth. The secondary

of the transformer consists of three stator windings, surrounding the primary like the windings of a three-phase generator. The transformer is constructed on pulse-transformer principles, so the trapezoid wave is passed through it with but minor deterioration in shape.

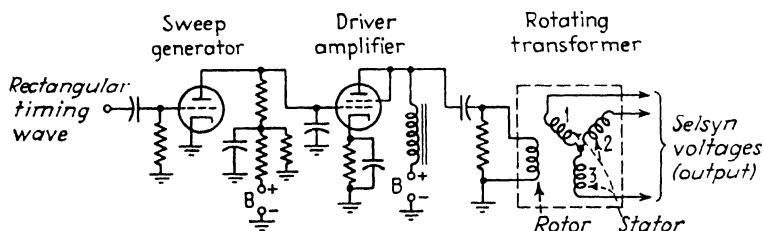



FIG. 411.—Circuit for imposing sinuosidal modulation by a three-winding rotary transformer.

Across the three secondary windings appear three sets of trapezoids, each modulated by a sinusoidal envelope. The three envelopes are displaced in phase from one another by 120 electrical degrees.

It is necessary to convert the three modulated trapezoids to four, in two groups, as shown in Fig. 412. The conversion is

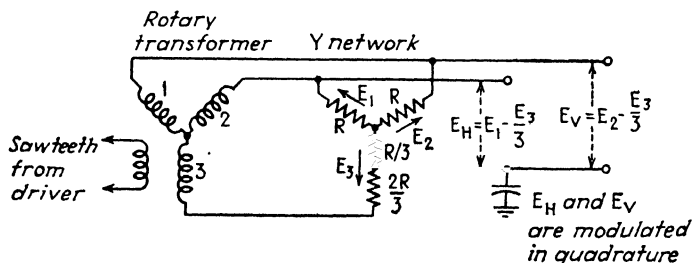



FIG. 412.—Circuit for converting three-phase to four-phase components.

accomplished by connecting the stator windings to a resistor network in the form of a Y. A tap is taken on the lower resistor element of the network, at a point one third down from the center junction. The trapezoids appearing between this point and the upper terminals, respectively, are modulated with envelopes 90 deg out of phase. Consequently two modulation envelopes in quadrature are obtained. The two sets are converted to push-pull relationship (making four sets) prior to application to the deflection coils.

When the modulated trapezoids of the fixed-yoke ppi system are passed through a capacitive or transformer connection, as is necessary in the subsequent deflection amplifier stages, the d-c component of the wave is replaced by the average of the waveform. Since the trapezoids are continually varying in amplitude

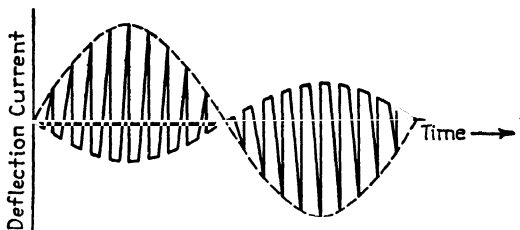



FIG. 413.—Distortion of deflection baseline when d-c component is removed.

in accordance with the modulation envelopes, the base of the waveform is not a straight line, but a sine curve, as shown in Fig. 413. If such a wave were applied to the deflection coils without reinserting the d-c component, the sweeps would start not from the center of the tube but from a variety of points displaced from the center. Consequently reinsertion of the d-c component

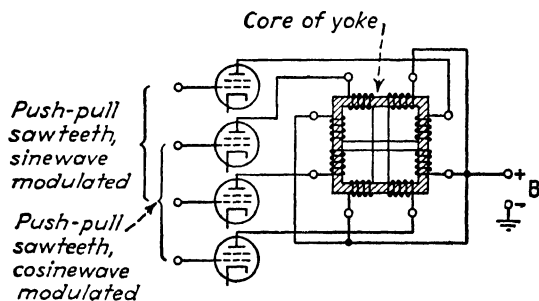



FIG. 414.—Output circuit (deflection amplifier and yoke) of fixed-yoke ppi deflection system.

is necessary. It is accomplished by clamping tubes connected to the grids of the deflection amplifiers and operated by the leading and trailing edges of a rectangular wave supplied by the timer.

The operation of such clamping circuits has been discussed in Sec. 144, in which a diode form is shown. In ppi applications, triode clampers are preferred since they are more positive in their action. A complete description of the d-c reinsertion system is given in Sec. 210 of this chapter.

The remaining portion of the fixed-yoke ppi system is the deflection amplifier. The scanning yoke has four separate coils, as shown in Fig. 414, each coil being wound on two legs of the core. The coils are wound in opposite directions so the d-c components of flux are opposed and there is no steady magnetization of the core. The two coils on opposite legs are fed trapezoidal voltage waves in opposed polarity (push-pull). The vertical and horizontal deflection systems are identical. Separate deflection amplifiers feed the coils, as shown in the diagram. The amplifiers are usually of the beam-power tetrode type, capable of supplying high current through small internal resistance.

206. Direction-sweep Potentiometers.—In most airborne radars the deflection currents for indication of azimuth and eleva-

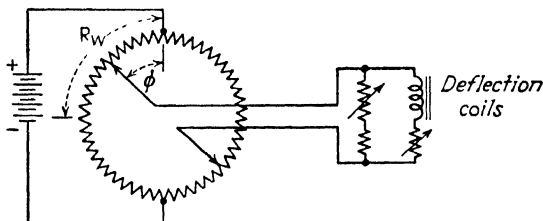



FIG. 415.—Circuit for generating direction-sweep deflection currents.

tion are derived from potentiometers (Fig. 415). The rotor element is connected, either directly or through gears and flexible shafting, to the corresponding shaft of the scanner. The stator element is a continuous winding of resistance wire on a cylindrical card. Two taps at opposite sides of the stator winding are connected to a current source, such as a storage battery. The rotor element has two brushes that bear on the winding at opposite sides, as shown. These brushes are connected to the deflection coil. As the contacts rotate, the current in the coil varies from zero (when the brushes are midway between the battery taps) to a maximum (when the brushes fall on the battery taps). At intermediate positions, the current is a linear function of the angle between the contacts and the taps. When the contacts rotate at constant speed, the current waveform is a symmetrical triangular wave.

Since the rate of the scanner rotation is not greater than several revolutions per second, the frequency of the deflection waveform is low, and the inductive reactance of the deflection coils is so

small that it can be neglected. However, the resistance of the deflection coils is appreciable, compared with the resistance of the potentiometer, and this fact must be taken into account in arriving at a linear relationship between deflection current and scanner angle. To achieve linearity it is necessary to incorporate a taper in the resistance winding. The form of the taper is derived as follows: Consider the potentiometer winding as made up of four quadrants, each having a resistance R_w. As the contact arm rotates from its reference position on the battery taps, the resistance increases as a function of the angular rotation, that is

$$R = A(\phi)R_w \tag{501}$$

where R is the resistance included between the tap and the brush, when the angle between them is ϕ and A is the function of ϕ (varying from 0 to 1) representing the required taper. To find the taper, we solve the circuit for the current in the coil I_1 as a function of the value of A, with the following result:

$$\frac{I_1}{I_{max}} = \frac{A}{1 + (R_w/R_1)(1 - A^2)} \tag{502}$$

where I_{max} is the maximum deflection current (which flows when the brushes are on the battery taps) and R_1 is the resistance of the deflection coil.

It is required that the load current I_1 be linearly related to the angle over the 90 deg range of each quadrant, so

$$\frac{I_1}{I_{max}} = \frac{\phi}{90°} \tag{503}$$

Substituting this relation in Eq. 502 we can solve for A as a function of ϕ, where R_w and R_1 are specified. Typical values of these resistances are $R_w = 100$ ohms per quadrant and $R_1 = 200$ ohms.

Solving Eq. (502) for these values gives

$$A^2 + \left(\frac{180°}{\phi}\right)A - 3 = 0 \tag{504}$$

For each given value of ϕ this quadratic may be solved for A. The resulting values of A run from zero to unity, as plotted in Fig. 416.

The taper is approximated in practice by three linear portions, shown in the figure. The card on which the wire is wound is notched so that the resistance per degree, of the wire wound on the notch, has the slope of the corresponding linear segment. To permit adjusting the deflection circuit when deflection coils are changed, it is usual to include a variable resistance across the coil which can be adjusted until the resistance presented has the value for which the potentiometer taper is designed.

To assure long life, the resistance wire is ground to a flat surface under the brushes, and the brushes themselves are wide enough to bear on two wires simultaneously. Several hundred hours operation is usually possible before the brushes must be replaced.

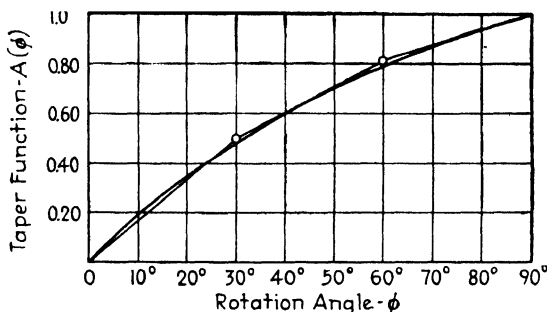



FIG. 416.—Theoretical taper for direction-sweep potentiometer (smooth curve) and straight-line approximation.

In many applications, deflection is required over a limited portion of the full rotation (say 90 deg or 180 deg). In this case cam-operated switches may be employed to disconnect the current source over the inactive portion of the sweep.

207. Data-transmission Systems.—The potentiometers just described are simple data-transmission systems that find use when an angular coordinate is to be transmitted to a cathode-ray indicator. Data-transmission systems are used for a variety of other purposes in radar, of which the following are outstanding:

1. Transmission of angular coordinates from scanner to a mechanical pointer (as a substitute for a cathode-ray indication).
2. Transmission of angular error signals (that is, signals related to the difference between a prescribed angle and the angle actually assumed by the scanner) to a mechanical indicator.

3. Transmission of angle and range coordinates from operator-controlled handwheels to associated gunfire- or searchlight-control mechanisms.

4. Automatic transmission of angle and/or range data from indicator circuits to the gunfire- or searchlight-control mechanisms.

While in theory any of the above functions might be performed by a potentiometer system, preference is accorded to another electrical transmission system on the score of reliability and accuracy. This is the selsyn (also widely known as "synchro") data-transmission system that we have already noted in one form in the fixed-yoke rotary-transformer ppi system (Sec. 205).

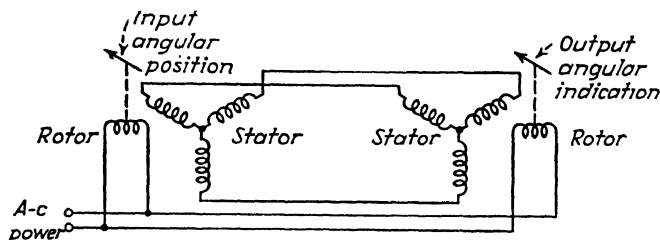



FIG. 417.—Selsyn transmitter and receiver.

Three forms of selsyn systems have found wide use in radar: the direct-transmission type, the differential type, and the control-transformer type. The direct-transmission selsyn is illustrated in Fig. 417. The system consists of two units, a transmitter and a receiver, which are electrically identical. Each consists of three stator windings, wound in slots in the stator core and displaced 120 deg from one another, and a rotary winding. The rotor and stator windings of the two units are connected together, as shown, and the rotors are fed alternating current from the same source.

When so energized, the rotor winding of the selsyn receiver takes up an angular position that corresponds to the position of the rotor in the selsyn transmitter. Thereafter any change in the position of the transmitter rotor is followed automatically by the receiver rotor, regardless of how irregular or intermittent the changes may be. The selsyn receiver unit is usually of sufficient size to drive directly an indicating shaft, rotary ppi-yoke, or similar device. Usually, however, selsyns are not large

enough to motivate a scanner directly, but do so through an intermediate mechanism.

The internal action of the selsyn is as follows: When the a-c power is applied to the transmitter rotor, it induces in the three stator windings alternating voltages displaced in phase 120 electrical degrees. These voltages are transferred without relative change in amplitude or phase to the corresponding stator windings in the receiver, so that relatively identical currents flow in the two sets of stator windings. The resultant magnetic field in the receiving unit is the vector sum of those produced by the three

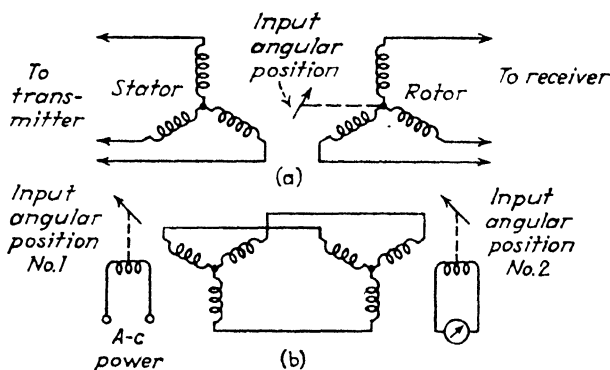



FIG. 418.—Differential selsyn (a) and control transformer (b).

windings, and, by the symmetry of the system, its direction is the same as that of the magnetic field existing in the transmitter unit. This latter field is in turn determined by the position of the transmitter rotor. The receiver rotor, being fed from the same alternating source, is constrained to take up the same position relative to the magnetic field in which it finds itself. Consequently, the two rotors take up corresponding positions and maintain them throughout all subsequent rotations imparted to the transmitter rotor. The two rotor positions may depart from coincidence momentarily, during an acceleration of the transmitter rotor, but this deviation is speedily corrected. When at rest the two positions agree within 1 deg, and some types are accurate to a small fraction of a degree.

The differential selsyn (Fig. 418*a*) is connected between the transmitter and receiver to permit introducing a second input rotation. The selsyn receiver then takes up a position that is the sum (or difference) of the two input rotations. The differ-

ential selsyn has three stator windings like the transmitter and receiver, and its rotor also has three windings. When like coils of rotor and stator are aligned, the current is transformed from one set of windings to the other without change of phase. But when the rotor is moved, the output phase is shifted by an amount equal to the rotation and this phase shift is passed to the selsyn motor, which responds accordingly. The differential selsyn system may be used with any two of its shafts acting as inputs, and the third taking a position equal to the sum or difference of the input rotations. Thus, for example, if the two input rotations are applied to the selsyn transmitter and receiver, the three-winding rotor of the differential unit will take up the differential position. It will be noted that the power consumed in the differential unit is provided by the stator windings of the transmitter and receiver. The latter units must accordingly be large enough to handle the additional power required.

When the connecting wires between units are long, as they may be when the controlling shaft (or shafts) and the controlled shaft are located in widely separate units of the equipment, appreciable losses may occur in the connecting leads. These losses must be taken into account in designing the selsyn system. The losses do not affect the accuracy of the indication, so long as all three currents are attenuated and delayed in the same proportion, but the available power at the output shaft is thereby reduced and the selsyn receiver may then prove incapable of turning the connected mechanical load.

The selsyn control transformer (Fig. 418*b*) is similar to the direct-transmission type except that alternating voltage is fed to the transmitter rotor only. An alternating voltage appears across the control transformer secondary (rotor) whose magnitude depends on the relative positions of the two rotors. The greater the difference in position, the greater the voltage, within the limits of plus or minus 60 deg. The voltage may be used to indicate the "error" (difference in angular position) between two input shafts connected to the two rotors. The output error voltage may be indicated directly on a meter or, as is more usual, it may actuate a control amplifier, which in turn operates a motor controlling one of the shafts. This comprises an automatic follow-up system or servomechanism, in which one shaft carrying a heavy load (say the shaft of the scanner) is

constrained to follow another shaft driven by a much smaller force, such as an operator's handwheel.

208. Servomechanisms and Power Drives.—A typical servomechanism, shown in Fig. 419, causes a scanner to take up a position corresponding to the position of a handwheel. A conventional direct-transmission selsyn system, shown at the left, indicates bearing in azimuth of the scanner by means of a pointer attached to the selsyn receiver. The servo system, at the right,

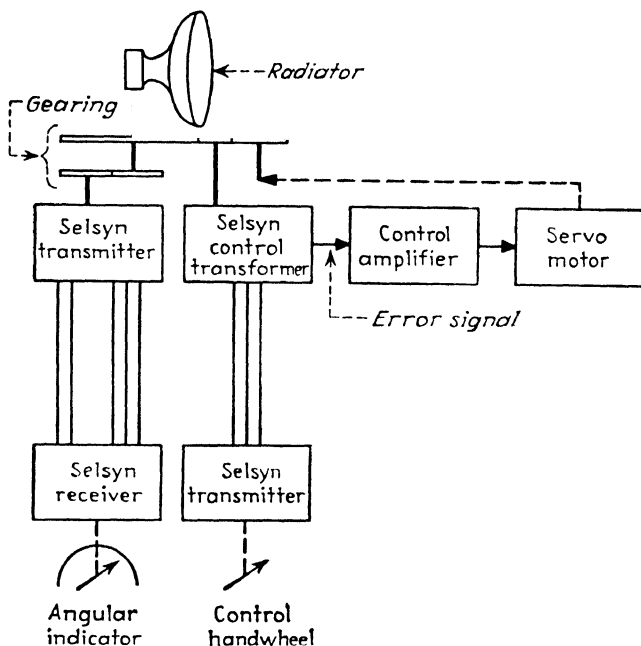



FIG. 419.—Elements of a servomechanism for control and indication of angular scanning.

starts with an azimuth control handwheel that controls the rotor of a selsyn transmitter. This is in turn connected electrically to a selsyn-control transformer that produces an output voltage ("error signal") proportional to the difference in the positions of the output and input shafts. The error signal feeds a control amplifier that in turn controls the servo motor, coupled to the scanner through gears.

When it is desired to turn the scanner, the handwheel is turned, producing an error signal which, operating through the control amplifier and motor, drives the scanner in the direction taken by

the handwheel. The scanner motion continues so long as the handwheel is turned, but when the handwheel comes to rest, the scanner continues to move only so long as there is a difference in their angular positions. This difference is reduced, as the scanner motion continues, until the error signal becomes zero and the system comes to rest at the "zero-error" position.

One difficulty in servomechanisms is the tendency to override the zero-error position, due to the inertia of the scanner. If this occurs an error signal of reverse sign appears and the motion of the scanner is reversed. The motion may then again override the zero-error position and a series of reversing motions called "hunting" occurs. If the constants of the oscillating system are

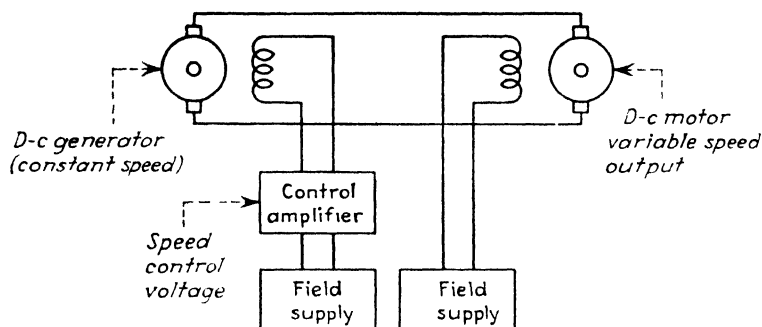



FIG. 420.—Ward-Leonard speed-control drive.

not properly chosen, the system may continue to oscillate about the zero-error position, and the scanner never comes to rest at the correct position. Even if the oscillations decrease in amplitude and the system eventually assumes the correct position, the consequent delay and the mechanical and electrical strain are serious disadvantages.

To avoid the tendency to hunt, it is necessary to reduce the driving force as the system approaches the zero-error position. This may be achieved by a feedback connection between drive motor and control amplifier, by the introduction of electrical damping in the control amplifier circuits, and by mechanical damping.

The basic elements of the servomechanism are the control amplifier and the motor drive. The two forms of drive most widely used are the Ward-Leonard drive and the amplidyne drive.

The Ward-Leonard drive (Fig. 420) consists essentially of a d-c motor which provides the output motive power of the unit and a d-c generator which supplies the armature current of the motor. The generator is driven at constant speed, and the direct current fed to the generator field is controlled by a d-c control amplifier. As the generator field current is increased, the generator output voltage increases, and the speed of the d-c motor increases. The system is usually driven by a-c power, an a-c motor driving the d-c generator and a rectifier (or d-c exciter generator driven by the a-c motor) supplying the d-c voltages required for the fields of the generator and d-c motor. By reversing the direction of the d-c generator field the direction of the motor drive can be reversed.

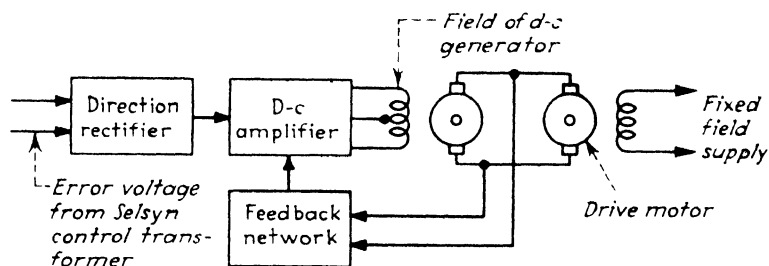



FIG. 421.—Control amplifier for Ward-Leonard drive.

The control amplifier for the Ward-Leonard drive (Fig. 421) takes its excitation from a selsyn-control transformer, which provides an error voltage proportional to the difference in shaft positions. The phase of the error voltage is compared with a reference voltage and a rectified direct voltage derived from it. The magnitude of this direct voltage is proportional to the magnitude of the position difference and its sign reverses as the output shaft position changes from behind to ahead of the input shaft position. The direct voltage controls a d-c amplifier through which the d-c generator gets its field supply. A feedback connection from the d-c generator output to the control amplifier reduces the gain in the amplifier as the voltage on the d-c motor decreases, that is, as the zero-error position is approached. In this way the tendency to hunt is minimized. The control function may also be performed mechanically through a series of switches that are energized in sequence by the rotation of a selsyn differential generator.

The amplidyne drive (Fig. 422) is similar to the Ward-Leonard except that a special form of d-c generator, with an extra set of short-circuited brushes, is used to drive the d-c motor. The advantages of the amplidyne system are its high power gain (output power up to 10,000 times the input power) and its freedom from electrical inertia (rapidity of response).

The short-circuited brushes are placed in the positions normally occupied by the load brushes. The short circuit passes a high current that induces a reaction flux at right angles to the normal excitation flux (produced by the field coil of the generator). This reaction flux is capable of inducing an output current across another set of brushes placed at right angles to the short-circuited set. This second set of brushes is connected to the load. The load current induces still another reaction flux that bucks the

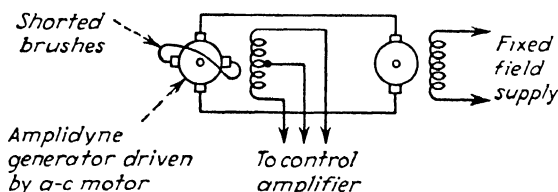



FIG. 422.—Amplidyne control system.

excitation flux. An additional compensating field winding, connected in series with the load, opposes this second reaction flux, so that the two remaining fluxes are the normal excitation flux and the reaction flux of the short-circuited brushes.

A d-c generator so constructed can generate a normal current output in its load with about 1 per cent of the field current normally required by a d-c machine. Since the field current and the associated excitation flux are small, a change in field current is very rapidly translated into a very much larger change in output current and voltage. Thus a servo system of high gain and rapid response is produced. The amplidyne generator is driven by an a-c motor, as in the Ward-Leonard system.

The field current of the amplidyne generator is obtained from the control amplifier shown in Fig. 423. This amplifier is essentially a combined push-pull and push-push a-c amplifier. The error voltage from the selsyn-control transformer is applied in push-pull to the grids of the amplifier, whereas a reference voltage of the same frequency is applied in push-push fashion

to the plates of the tubes. When the input error voltage is zero, the reference voltage causes equal currents to flow in both tubes; and therefore equal currents flow in the amplidyne generator field coils, which are arranged to buck one another. Consequently the amplidyne receives no field excitation and no output appears, that is, the system remains at rest.

When the error voltage has a finite value due to angular displacement between input and output shafts of the selsyn system, the currents in the two tubes are unequal by an amount depending on the relative magnitude and phase of the error voltage with respect to the reference voltage. Consequently the currents in the opposed portions of the amplidyne field coils are not equal,

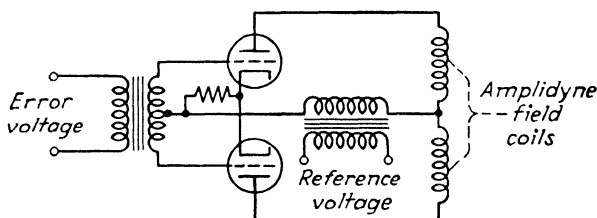



FIG. 423.—Control amplifier for amplidyne drive.

and a net flux appears in the generator, the amount and direction of the flux corresponding to the magnitude and phase of the error voltage. The amplidyne then produces an output current that drives the d-c servo motor in the corresponding direction. The shaft of the motor is connected, directly or through gearing, to the scanner shaft and the selsyn-control transformer. The direction of the motion is such as to reduce the error signal and thus bring the system to rest with input and output shafts at the same angle. Hunting is minimized by feeding the amplidyne generator output voltage (through an *RC* damping network and voltage divider) back to the input of the control amplifier, in degenerative polarity.

The application of the servomechanisms just described is by no means limited to the translation of a handwheel position into a corresponding scanner position. One application of particular value in shipborne installations is the indication of relative bearing in azimuth. As the ship maneuvers in azimuth, the indicated target azimuth appears to change. To stabilize the azimuth indications it is customary to introduce a shift in the indicated

azimuth by connecting a differential selsyn system to the gyro compass of the ship. The indicated azimuth is then compounded of two motions, one due to the target and the other due to the ship. The target azimuth is thus always referred to true north, regardless of the ship's heading.

209. Typical Indicator and Scanner Systems. SCR-268.—As a first example of scanner and indicator systems, we choose

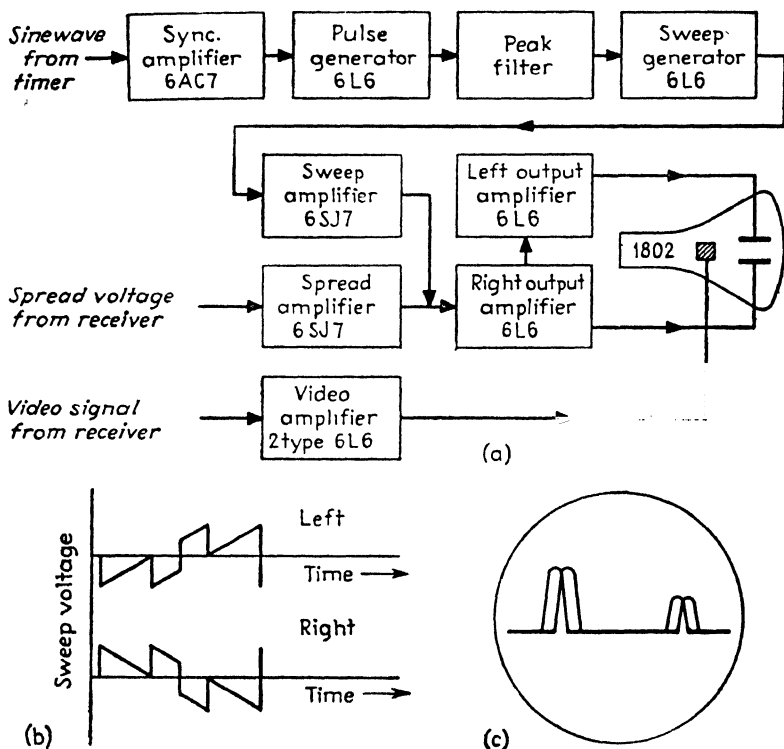



FIG. 424.—Block diagram of SCR-268 indicator (*a*), waveform (*b*), and presentation (*c*).

the simple (and largely outmoded) equipment of the SCR-268 searchlight-control radar. As we recall from Sec. 180, the timing of the SCR-268 is determined by a sine wave at pulse-rate frequency (4,098 pps). The sine wave is generated in an LC oscillator and passed through a succession of limiting and filtering stages to a Helmholtz-coil phase shifter (see Fig. 324). The output of the phase shifter is amplified and applied to the input of the indicator proper. Three separate indicators are used in the

SCR-268, one for indicating range only and the other two for matching the lobe-switched signals in azimuth and elevation, respectively. The system for indicating the azimuth angle is shown in block diagram in Fig. 424. The phase-shifted sine wave from the range unit (Fig. 324) is applied to a synchronizing amplifier, which narrows the negative portions of the sine wave. This wave drives a beam-power pulse generating tube that produces nearly rectangular waves. These are passed through a differentiator and the peaked waves control a sweep generator tube that produces a sawtooth wave. The phase of this wave is reversed in passing through the sweep amplifier stage.

The sweep voltage thus produced would be applied (after amplification) to the deflection plates of the type A indicator tube were it not for the fact that two separate signals, resulting from lobe switching, must be displayed simultaneously. The separation of the two signals is accomplished by shifting the sweep voltage bodily left or right by the superposition of a rectangular wave ("spread voltage") derived from the lobe-switching oscillator in the receiver (see Fig. 376). The combined sawtooth and rectangular wave is shown in the Fig. 424. The combined sweep wave is then passed through two beam-power deflection amplifiers that reproduce the deflection voltage in push-pull form, the two push-pull components being applied to opposite horizontal deflection plates of the indicators. A 5-in. non-persistent indicator tube is used.

The signal indication is applied to the vertical deflection plate from the receiver output. The transmitted pulse and echo pulses thus appear as in the type A indication, and each pulse is split into two images, one corresponding to reception from the left-hand lobe of the antenna, the other to reception from the right-hand lobe. The sweep lasts for nearly the full pulse interval of 244 μsec, so that echoes out to a maximum range of 22 miles may be accommodated.

In operation, the phase shifter is adjusted until the echo of particular interest lies under a hairline at the center of the indicator screen. The azimuth operator then turns a handwheel that drives the entire antenna assembly, operators and all (see Fig. 340), in azimuth until the two aspects of the split image have the same amplitude. An identical scanning system, operated by the elevation operator, turns the antenna system in elevation until

the split image on the elevation indicator is equalized. As the target moves, the azimuth and elevation operators continually turn the wheels to maintain equal amplitudes in the split image. Simultaneously the range operator, who operates the phase shifter, keeps the desired echo in the center of the screen.

The rotations of azimuth and elevation handwheels and range of the phase shifter are transmitted by selsyn data-transmission systems to the associated searchlight-control mechanism. The SCR-268 was originally designed for gunfire control also, but its accuracy was hardly sufficient for this purpose and this function was taken over by more modern equipment, notably the SCR-584 and AN/MPG-1 radars.

210. Scanner and Indicator of the AN/TPS-3 Radar.—The timing circuits of the AN/TPS-3 600-megacycle early-warning radar have been described in Sec. 175. This radar has two indicators, type A and ppi. The sawtooth generators for each of these indicators have already been described (Fig. 315). The remaining portion of the indicating system is the ppi proper, shown in Fig. 425. This circuit is of the fixed-yoke rotary-transformer type. The rotary transformer is mounted in the antenna pedestal. The rotor of the transformer moves with the radiator and is fed sawtooth waves through slip rings at a peak value of about 35 volts.

The stator of the rotary transformer consists of four windings, which are arranged 90 deg apart and are distributed in slots in the stator core. The sawtooth voltage applied to the rotor is induced in the four wings as four separate sawtooth sequences, each modulated with a sinusoidal envelope displaced 90 deg from its neighbors. The ends of opposite coils are connected internally and four leads are taken from the remaining terminations. Between these leads and ground the sinusoidally modulated sawteeth appear. The windings are capacitively coupled to four deflection amplifier tubes of the beam-power type.

To assure reinsertion of the d-c component, a clamping circuit of two triode tubes is connected conductively to the grid of each deflection amplifier tube (to avoid confusion only one clamping circuit is shown). The clamping tubes are connected in series, that is, the plate of the lower tube is connected to the cathode of the upper tube. The plate of the upper tube is connected to a 90-volt B supply. The grids of both tubes are connected through

FIG. 425.—Simplified schematic of AN/TPS-3 ppi indicator.

a resistance to the same B supply voltage and in consequence both tubes normally conduct. The two tubes thus form a voltage divider. The deflection amplifier grid, connected between them, is held rigidly at a potential determined by the voltage drops through the tubes and the cathode bias resistance of the lower clamping tube. The clamping tubes constitute a very low impedance across the deflection amplifier grid and thus prevent any deflection voltage that may be present from affecting the output current of the deflection amplifier.

The grids of the clamping tubes are connected capacitively to the output of a multivibrator that operates synchronously with the sawtooth deflection. This multivibrator is, in fact, the same circuit that drives the sawtooth generator (Fig. 315). A negative sawtooth wave of about −160 volts amplitude is produced by the multivibrator, the leading edge of the wave occurring at the instant the sweep starts. This negative wave, applied to the grids of the clamping tubes, cuts them off sharply at the beginning of the sweep and thus allows the deflection amplifier grids to follow the sawtooth wave that appears at the same instant. Thus all the deflection amplifier tubes start to follow the sawtooth at the same instant from the same value of grid voltage, and the deflection is constrained to start from a fixed position which is adjusted to fall at the center of the screen. At the end of the sweep the negative wave ceases and the clampers resume control of the amplifier grids.

The outputs of the deflection amplifiers feed the deflection yoke, which consists of two center-tapped coils wound in eight symmetrical portions on the yoke core. Each coil is loaded with shunt resistance and the center-taps are returned to the 300-volt B supply. The grids of the lower two deflection amplifiers are shunted by an adjustable resistance that adjusts the peak amplitude of horizontal sawteeth relative to the vertical ones and thus permits adjusting the system until the sweep rotates uniformly in synchronism with the radiator.

The intensity modulation of the ppi indicator consists of the echo signals from the receiver and the marker pulses from the timer (see Fig. 426). These are both of positive polarity and are applied to the control electrode of the ppi tube.

The scanner of the AN/TPS-3 is of simple construction. It consists of a vertical member to which the radiator is attached,

guyed at the top and attached at the base to a standard which protrudes through the tent which houses the crew. At the base is a motor drive that rotates the radiator in azimuth at a rate of about 10 rpm. This motor operates under the control of a two-way reversing switch, so that the radiator may be rotated continuously or intermittently, in either direction. The pedestal also contains the rotary transformer of the ppi and terminations for the leads of the rotor and four stator windings. The radiator itself is a 10-ft paraboloid of wire mesh fed by

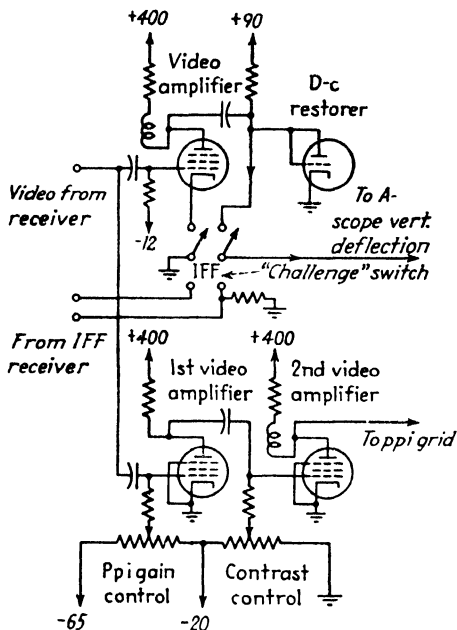



FIG. 426.—Signal circuits of AN/TPS-3 indicators (type *A* and ppi).

a multiple dipole feed that has two alternative connections (cf. Sec. 186 and Fig. 343). The radiator may be located 100 ft from the tent housing the equipment, in the interest of protecting the crew from enemy gunfire. Connecting cables carry the motor-drive power, the ppi rotary-transformer connections, and convey the r-f pulses to and from the radiator.

211. Scanner and Indicator of the SCR-584.—We turn now to the indicating and scanning portions of the SCR-584 gunfire-control radar. The timing functions of this equipment, including the sweeps for the two type J range indicators, have already been

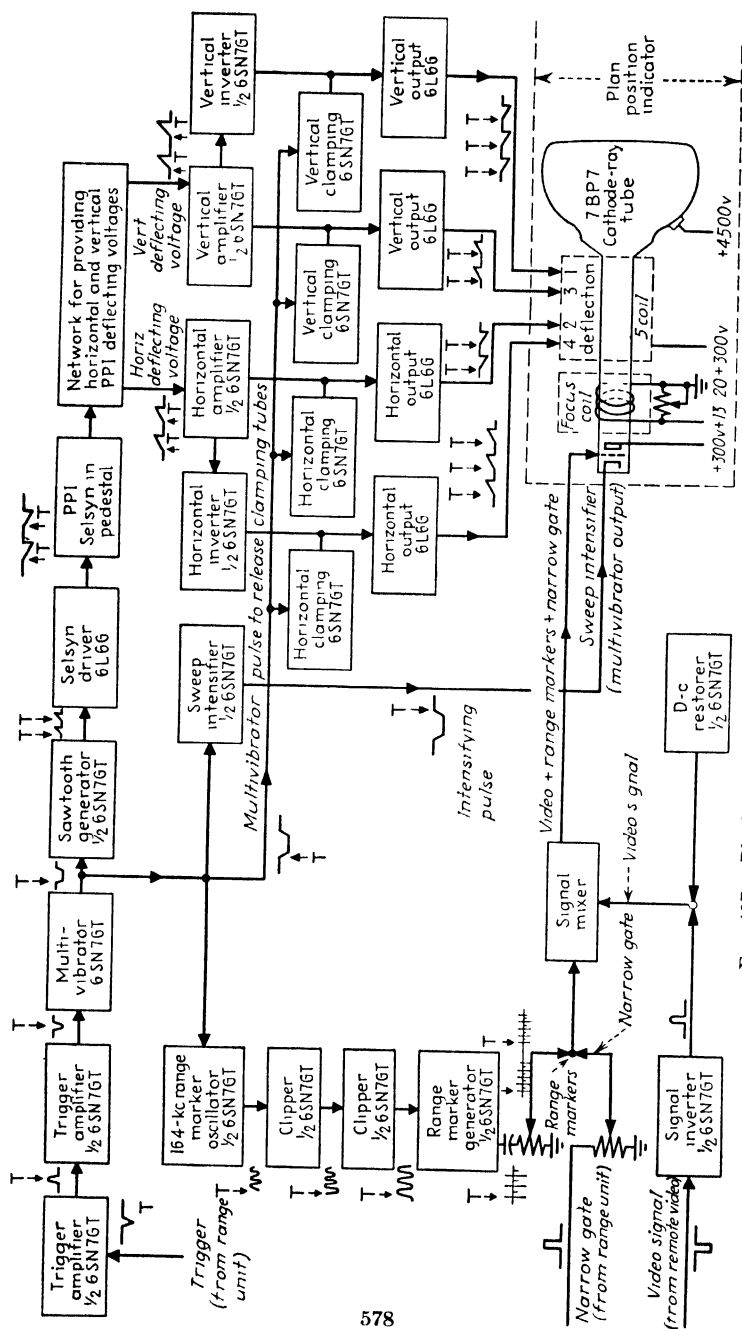



Fig 427 —Block diagram of SCR-584 plan-position indicator.

described (see Sec. 178 and Fig. 322). The remaining portion of the indicating system, the ppi, is illustrated in block diagram of Fig. 427. We examine first the circuits producing the rotating sweep, which are of the fixed-yoke rotary-transformer type. The indicator accepts at the upper left corner of the diagram a trigger from the timer, which, we recall from Sec. 177, occurs coincidentally with the transmitted pulse. The trigger is first amplified in two stages and then applied to a multivibrator that produces a negative rectangular wave of adjustable width. A switch selects either of two widths, one corresponding to 35,000 yd maximum range, the other to 70,000 yd. The leading and trailing edges of the multivibrator wave control a sawtooth current generator (trapezoidal voltage waveform). The output of this circuit is a succession of sawteeth at constant amplitude, synchronous with the initial input trigger.

The sawteeth are modulated by passage through a rotary transformer of the three-winding selsyn type (cf. Sec. 205, Fig. 411). Sinusoidally modulated sawtooth waves are induced in each of the three windings, the envelopes being displaced 120 deg in phase. The three modulated sequences are converted to two sequences, separated 90 deg in phase, by passage through a tapped Y resistor network (cf. Fig. 412). The horizontal and vertical deflection voltages thus obtained are converted to push-pull form by passage through an inverter-amplifier stage and thence passed to the four deflection amplifiers. At the grid of each amplifier is connected, conductively, a double-triode clamping circuit, identical in principle (though somewhat different in operating voltages) to that shown in Fig. 425. The grids of the clamping tubes are cut off during the sweep by the rectangular wave output of the multivibrator previously mentioned. The deflection amplifier outputs feed the deflection yoke, which consists of two center-tapped coils wound in eight portions symmetrically on the core. The center taps are connected to the 300 volt B supply. The indicator tube proper is a 7-in. long-persistence unit, type 7BP7. The tube is magnetically focused and the second-anode voltage is 4,500 volts.

The three intensity-modulation signals applied to the indicator are the target echo, the range marker pulses, and the intensifier pulses. The intensifier is a negative rectangular wave derived from the multivibrator and passed through a cathode-coupled

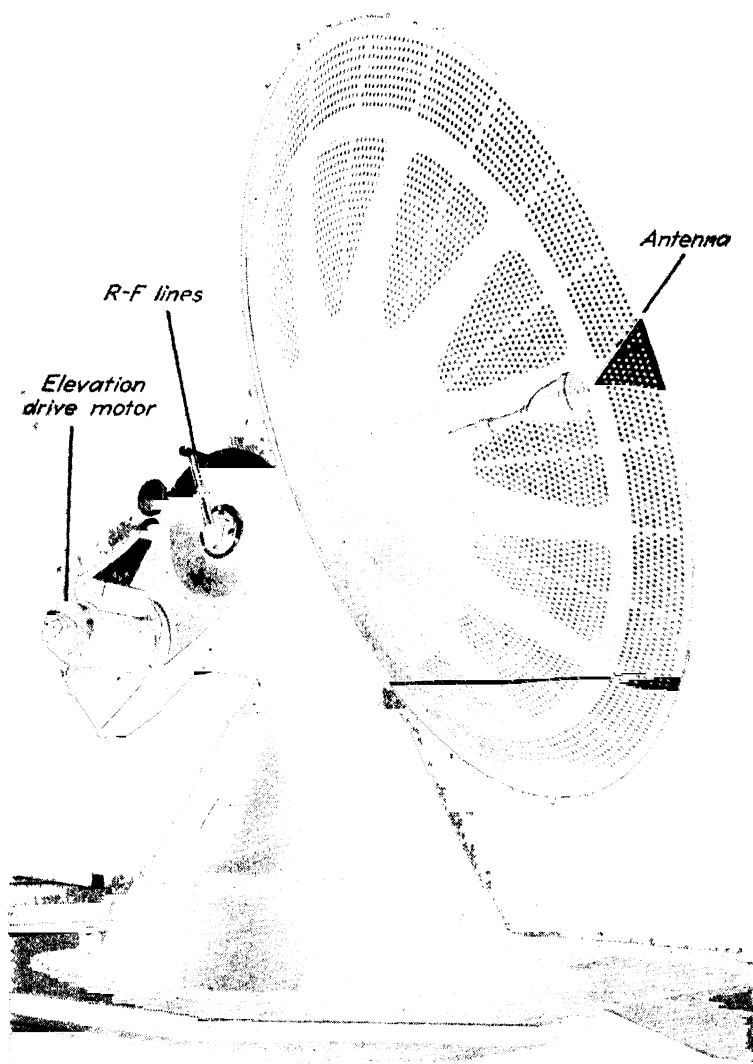



FIG. 428.—SCR-584 scanning mechanisms, showing rotating feed for conical scanning and drive for helical scanning.

stage. This negative wave, applied to the cathode of the indicator tube, brightens the beam for the duration of the sweep (70,000 yd or 35,000 yd).

The range markers are produced by a normally cut-off stage with a 16.4 kc tuned circuit in its cathode. The multivibrator negative wave cuts off this tube and excites damped oscillations in the tuned circuit. The period of these oscillations, 61 μsec, corresponds to a range interval of 10,000 yd. The damped sinusoids are clipped in two successive stages, forming rectangular waves of nearly constant amplitude. Passage through a differentiator circuit develops sharp pulses from the leading and trailing edge of the wave. The positive peaks serve as the range markers.

To these fixed range markers is added an adjustable range marker, the narrow gate previously described (Sec. 177). This gate is produced in the timer for the purpose of intensifying a portion of the vernier range indicator (type J, Fig. 321), on which the target echo of interest is displayed. The narrow gate produces a concentric circle on the ppi screen whose radius corresponds to the range then observed on the fine type J indicator. In operation, the range handwheel is adjusted by the operator until the movable range marker falls on the target of interest. The echo of that target is then displayed on the fine type J range indicator and lies under the airline on the coarse type J range indicator. The narrow gate and the range marker pulses are combined in a signal mixer stage.

The target echoes, from the receiver output, are amplified in the indicator unit and their d-c component restored. The echo pulses are mixed with the narrow gate and the range markers in the signal mixer previously mentioned. The combined intensity-modulation signal is passed to the control electrode of the indicator tube. Manual control of brightness is provided by a potentiometer in the cathode circuit.

The scanner of the SCR-584 is illustrated in Fig. 428. The r-f portions of this structure have already been discussed (Sec. 187). The scanner itself consists of two amplidyne motor drives for azimuth and elevation motions. These drives are motivated by automatic tracking circuits, the block diagram of which is shown in Fig. 429.

When the narrow-gate signal has been set over the target of

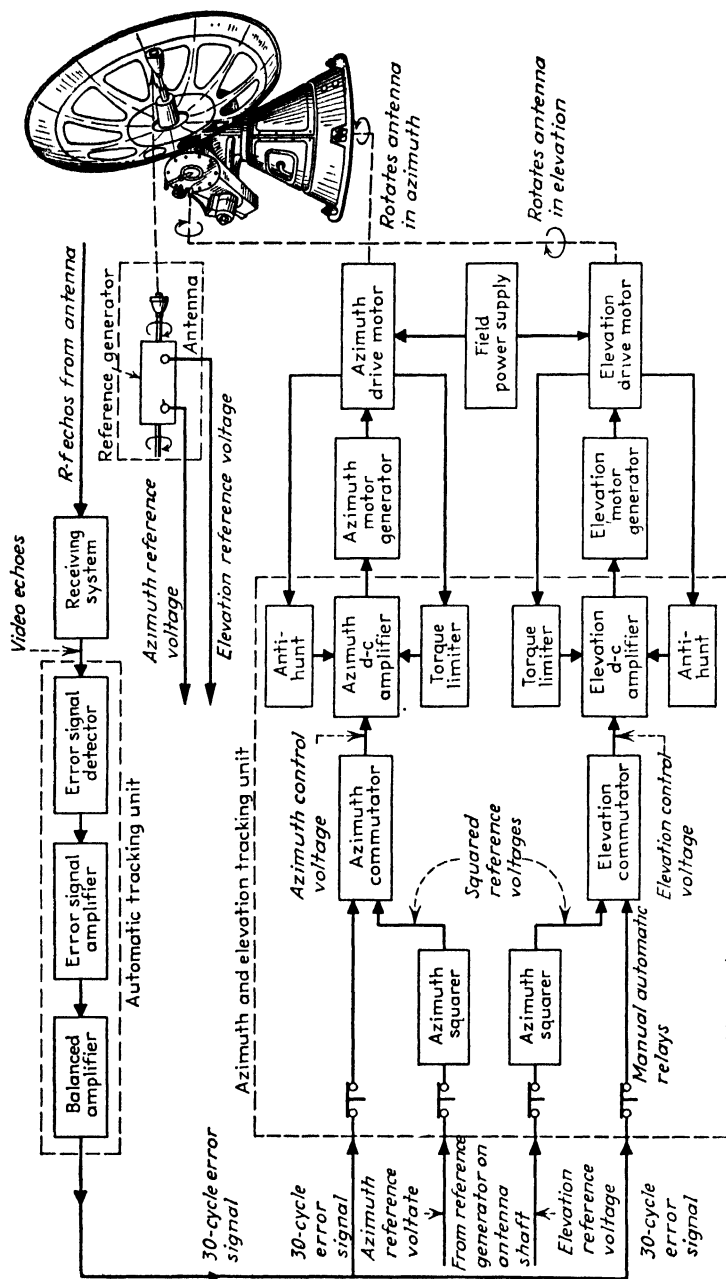



FIG. 429.—Block diagram of SCR-584 automatic tracking system.

interest, its echo appears on the range scopes. The system is switched to the automatic tracking position, and the following events take place: The narrow gate is applied in positive polarity to the grid of one of the i-f amplifier stages in the servo channel (Fig. 378). This permits the servo i-f channel to pass the selected target echo to the automatic control circuits, while blocking all other signals.

The selected echo sequence is modulated in amplitude, under the effect of conical scanning, whenever the target bearing lies off the axis of the conical scan. The sinusoidal modulation ("error signal") is recovered by passage through a third detector and converted to push-pull form. The push-pull components of the error signal are then compared with a reference alternating voltage of the same frequency and constant phase. This reference voltage is obtained from an alternator mounted directly on the shaft of the conical scanner. This alternator produces two voltages in quadrature, at the frequency of the conical scanner (1,800 rpm, 30 cps).

Each reference voltage from the alternator is squared and applied to the plate circuit of the corresponding amplidyne-control amplifier. Consider first the azimuth control. The corresponding (azimuth) error signal components are applied to the grids of the control amplifier. The control amplifier output (cf. Fig. 423) is zero when the azimuth error signal is zero and assumes a d-c component when an azimuth error signal is present, the magnitude and sign of the component depending on the deviation and direction of the target bearing in azimuth relative to the azimuth reference. The d-c component is averaged in a filter and applied to a d-c amplifier that develops the field current to the azimuth amplidyne generator. The generator in turn controls the azimuth drive motor, which turns the scanner in azimuth in the direction to reduce the azimuth error voltage. An antihunt connection from the drive motor terminals to the d-c amplifier suppresses hunting about the zero error position. An identical system operates to reduce the elevation component of the error signal.

Auxiliary selsyn data-transmission systems deliver the azimuth and elevation coordinates to mechanical pointers facing the operators so they are always apprised of the direction of the target. The same selsyn system transmits the data to the gun-

fire computer that introduces the necessary lead angles and controls the direction of the associated guns.

The range coordinate is not tracked automatically in the SCR-584 (such automatic range circuits are incorporated in the SCR-545 radar). In the SCR-584, the operator turns a handwheel that determines the time of occurrence of the narrow gate. The narrow gate intensifies a portion of the fine range indicator, so by turning the wheel it is possible to keep the selected target echo centered in the brightened portion of the sweep. The position of the narrow gate then corresponds to the position of the target, to a dynamic accuracy of 15 to 25 yd. The handwheel is geared to a selsyn-transmission system that feeds the range information to the gunfire computer.

A useful auxiliary in range tracking is the aided tracking system. When this device is used, the range handwheel is turned by a motor drive at an adjustable rate. If the target happens to be approaching or receding from the radar at constant velocity, the operator is relieved of the task of following the target echo. If the rate of approach or recession changes, the echo signal begins to drift out of the illuminated portion of the fine range scope. The operator, noting the drift, turns an auxiliary handwheel that performs two functions: it introduces a shift in position of the narrow gate sufficient to bring the sweep back to the target, and simultaneously it changes the rate of the automatic drive to make it more nearly approach the changed rate of approach of the target. By successive motions of the aided tracking handwheel, the operator is able to follow the target smoothly and accurately.

212. Scanner and Indicator of the AN/MPG-1 Radar.—The scanner of the AN/MPG-1 coastal-gunfire-control radar has already been described (Sec. 162 and Fig. 292). The scanner has two functions: searching for targets and tracking them during gunfire. In the search function, one of the arms in the rapid scanner is brought to rest at the center of the throat of the folded horn (Fig. 292) and the entire radiator is rotated in azimuth at a rate of about 5 rpm. The pulse rate is 1,024 pps. The indicator system then used is the ppi. The indicator is a 7-in. long-persistence tube, fitted with a rotating yoke driven by a selsyn system in synchronism with the scanner motion in azimuth. Two maximum ranges, 80,000 and 30,000 yd, are

provided. The 80,000-yd display includes eight 10,000-yd fixed range markers. The 30,000-yd range has a movable marker that is adjusted by the operator until it falls over the target of particular interest. The ppi system (Fig. 430) does not differ in any major essential from those previously described (except in its employment of the rotating yoke) and consequently will not be described in detail.

When the variable range marker is placed over the selected target, the radar is switched to the tracking function. The pulse

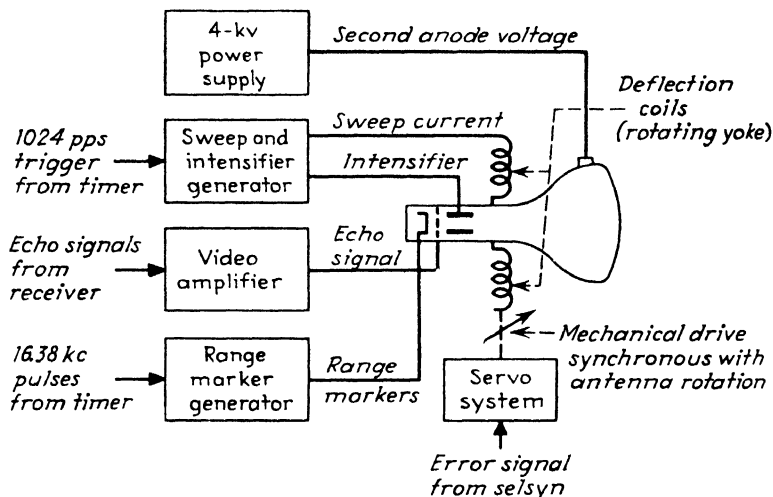



FIG. 430.—Block diagram of AN/MPG-1 ppi indicator.

rate is thereby stepped up to 4,097 pps, the spoke feed of the high-speed scanner goes into motion, and the ppi indicator is turned off. In its place an expanded type B indicator is used. This indicator (Fig. 431) covers a segment of 2,000 yd in range (vertically) by 10 deg maximum in azimuth (horizontally). Range and azimuth markers are imposed on the display at 1,000-yd and 1-deg intervals, respectively. The display on the type B indicator is the so-called "distortionless" or constant-scale presentation, in which the horizontal azimuth deflection is fixed at 400 yd per in. (equal to the vertical range scale), so that all targets are shown in fixed relation to each other, regardless of the range at which they are viewed. This feature is useful in one the principal functions of the type B indicator,

viewing the splashes of the shells relative to the position of the target and correcting the aim of the gunfire accordingly.

The block diagram of the type B indicator system is shown in Fig. 432. The timing function begins with a 163.9-kc crystal-

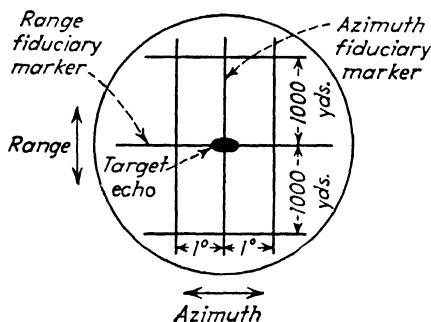



FIG. 431.—Expanded type B indicator of AN/MPG-1 radar.

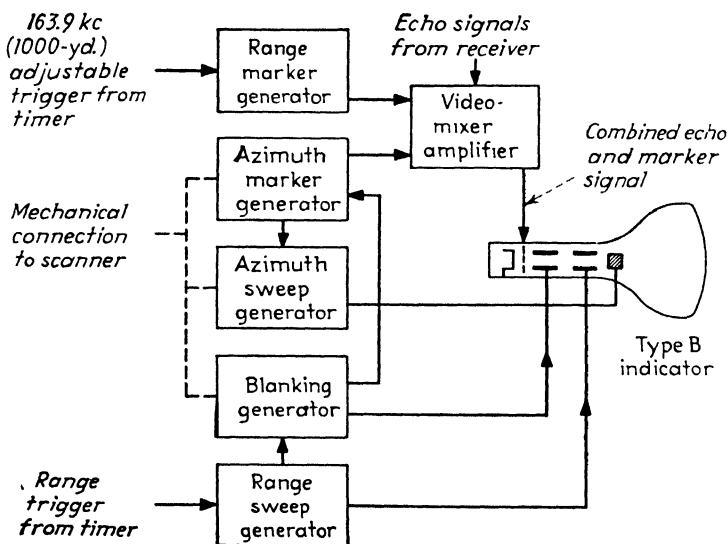



FIG. 432.—Indicator and auxiliary marker circuits of AN/MPG-1 type B indicator.

controlled sine-wave oscillator, whose period corresponds to a range interval of 1,000 yd. The output of this oscillator is passed through a quadrant-type capacitive phase shifter, and thence to a peaking amplifier that develops 1,000-yd range marker pips from the phase-shifted sine waves. This sequence of range markers is then passed through a selection circuit (coincidence

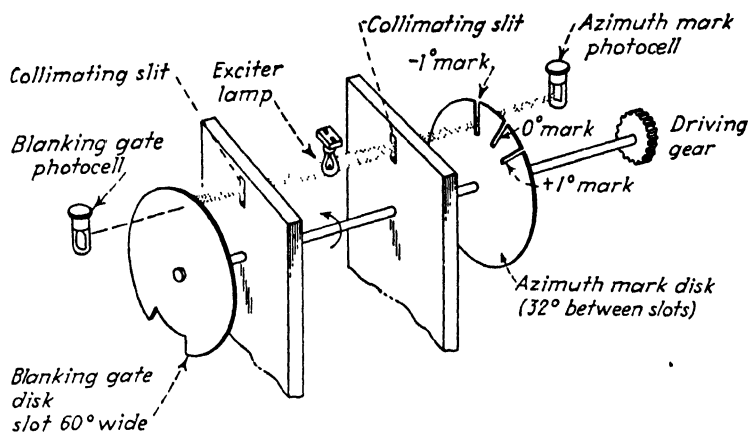



FIG. 433.—Electro-optical system for generating azimuth markers in AN/MPG-1.

Antenna beam angle
-5°
-1°
0
+1
+5
Useful antenna output
Rotating feed arm output
Feed arm (#1) energized
Feed arm (#2) energized
Feed arm (#3) energized)
Feed arm (#4) energized
Feed arm (#1) energized)
Azimuth sweep voltage
Azimuth marks
Azimuth blanking gate
0.055 second
0.062 second
(one feed arm sweep)
(one rotation azimuth sweep mark unit)
0.25 second
(one rotation feed assembly)
Time

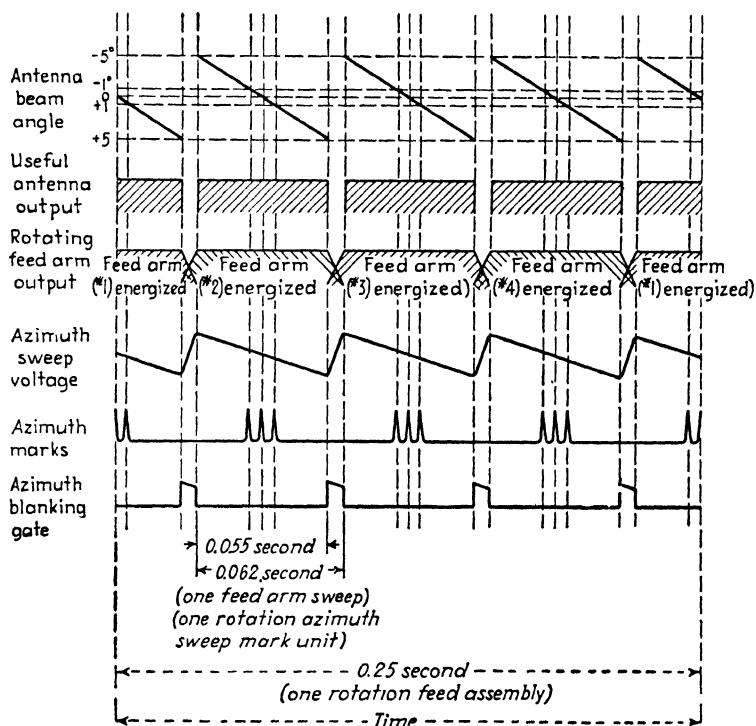



FIG. 434.—Time relationships of angle markers in AN/MPG-1.

amplifier) activated by a 6-μsec gate pulse in synchronism with the pulse rate, 4,097 pps. The position of this gate is determined by a range-delay circuit geared to the handwheel. The 6-μsec gate and one of the 1,000-yard range markers appear coincidentally and are passed through the coincidence amplifier. The output pulse of this amplifier triggers the range sweep of the type B indicator and forms the movable range marker on the

FIG. 435.—Ppi view of Pearl Harbor on AN/MPG-1 (30,000 yard sweep). The circle is the movable range marker set at 12,900 yards. It intersects the target echo from a string of LCI landing craft (*bottom*).

ppi. Consequently the range segment shown on the type B indicator and the movable range marker are synchronous. The three following 1,000-yd range markers appear on the type B indicator.

The azimuth markers are produced by the electro-optical system shown in Fig. 433. Two disks, driven by gears coupled to the spoke feed of the high-speed scanner, admit light from a lamp to each of two phototubes. Narrow slits in one disk correspond to the dead-ahead position of the radiated beam and to positions 1 deg either side. The photocurrent is amplified, peaked, and applied to the type B indicator to produce the azimuth markers. The other disk in the system provides

blanking voltage that extinguishes the beam during the azimuth sweep retrace. The timing coordination between the rapid scanner motion, azimuth sweep voltage, markers, and blanking is shown in Fig. 434. The azimuth sawtooth sweep is generated by a rotating capacitor that modulates a 1-megacycle oscillator with a sawtooth envelope. This envelope is recovered in a detector, amplified, and applied to the horizontal deflection system of the type B scope.

The distortionless type B presentation (400 yd per in. in both dimensions at the center of the scope) is achieved by expanding the azimuth sweep in conformance with the setting of the range delay circuit. As the range segment covered by the B indicator is moved outward, the 1-deg azimuth marks move further apart. Beyond 12,000 yd range, in fact, the full 10-deg segment in azimuth cannot be contained within the face of the indicator. However, the 1-deg azimuth markers still remain visible out to the maximum tracking range of 28,000 yd. Photographs of the ppi and type B presentations are shown in Figs. 435 and 436.

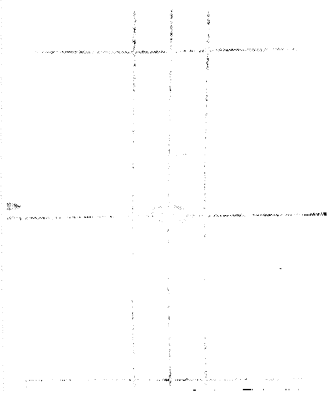



FIG. 436.—Type B indication from AN/MPG-1, showing target raft at intersection of fiduciary marks (*center*) and shell splash (*above*). The shell has landed about 400 yards over and about 0.3 degree to the right of the target.

The AN/MPG-1 radar embodies many of the more modern developments in scanner and indicator practice. Space does not permit a more detailed examination of its circuits. Additional information may be obtained from the reference cited below.[1]

[1] STRAUS, RUEGER, WERT, REISMAN, TAYLOR, DAVIS, and TAYLOR, The AN/MPG-1 Radar, *Electronics*, **18 (12)** 92, **19(1)** 110; **19(3)** 140 (December, 1945; January and March 1946).

CHAPTER XII

RADIO-FREQUENCY MEASUREMENTS AND TEST EQUIPMENT

Measurements in radar are concerned with either of two aspects of the system: the pulse aspect or the radio-frequency aspect. The pulse aspect has much in common with other branches of the radio art, notably television, and the measurement methods employed are not new. But in the r-f parts of the system another situation obtains. Radar was the first application of radio to employ ultrahigh and superhigh frequencies on a large scale. Coincidentally with the growth of radar, and necessary to it at each stage of its development, have appeared many new r-f measurement techniques. This chapter considers some of the more important of these r-f measurements and describes the equipment by which they are carried out.

213. Ultrahigh-frequency and Superhigh-frequency Measurements.—Since circuit elements in the uhf (ultrahigh-frequency) and shf (superhigh-frequency) regions are distributed over volumes or surfaces, the familiar quantities of voltage and current are not readily identified. In their absence, it is necessary to measure power and impedance in some way that does not involve the product or ratio of voltage and current. Similarly, although frequency is measured by resonant circuits, as in lower frequency practice, the resonant circuit elements are distributed and cannot be isolated from the resonant system. Consequently, the measurement of stored energy (Q, circuit merit) is performed, not by measuring reactive and resistive elements, but by noting the width of the resonance curve at a specified power level.

The four quantities met in radar r-f measurements are power, frequency, impedance, and stored energy. Power is generally measured by thermal means, which implies that the power actually measured is averaged over a time determined by the thermal capacity of the measuring element. For testing the power of a pulsed source (the radar transmitter, for example),

the average power is measured and the peak power computed from the pulse waveform, as revealed on an oscilloscope. In routine testing of circuit elements, a c-w (continuous-wave) test oscillator ("bench oscillator") is customarily used. The power itself is measured by a sensitive bolometer at low-power levels, or by a water calorimeter at higher levels. In the measurement of power, as in other quantities, the test equipment must include tuners and couplers to assure proper transfer of energy from one element of the system to another. At high levels, a substantial terminating load is required to absorb the power. Fixed and variable calibrated attenuators are also useful in power measurements.

Frequency is measured by reference to primary or secondary standards. The primary standard is generally a low-frequency (50 to 100 kc) temperature-controlled quartz crystal with auxiliary circuits to multiply the frequency to the uhf and shf ranges. The secondary standards are calibrated wave-meters of the coaxial or cavity types.

Impedance and admittance in the uhf and shf ranges are measured by standing waves. The standing-wave detector ("standing-wave machine") consists of an adjustable probe inserted into a slot along the coaxial line or waveguide carrying the r-f energy. A power-measuring bolometer or crystal, attached to the probe, indicates the positions and magnitudes of the maxima and minima in the standing-wave pattern. From these data, plotted on the impedance or admittance circle diagram, the normalized impedance of the termination can be computed. The standing-wave detector is also used in adjusting terminations and circuit elements to the "matched" condition, representing maximum transfer of power.

Stored energy is computed from the resonance characteristic of the device whose Q is in question. This measurement may be carried out almost instantaneously by the spectrum analyzer, which applies a variable frequency to the device under test and indicates the response, against frequency, on the screen of an oscilloscope.

214. Bench Oscillators.—Oscillators used as sources of test power at frequencies from 500 to 2,500 megacycles generally employ some form of disk-seal tube as the oscillator. At frequencies above 2,500 megacycles, klystrons are generally

used. The klystron, in one of the many available forms, can supply power up to about 40 watts at the lower end of the range (2,500 to 3,000 megacycles) and up to 5 watts at the higher end (10,000 megacycles). The reflex klystron is widely used because it is easily tuned by changing the voltage applied to the repeller electrode. This is an experimental convenience when the adjustment is performed manually, and a necessity in f-m (frequency-

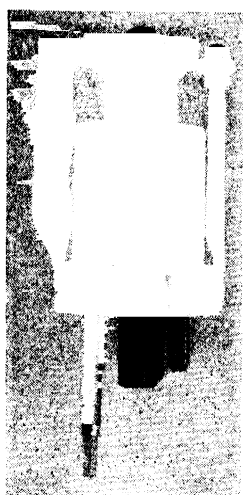



FIG. 437*a*.—External appearance of type 723A/B (2K25) local oscillator 3-cm reflex klystron.

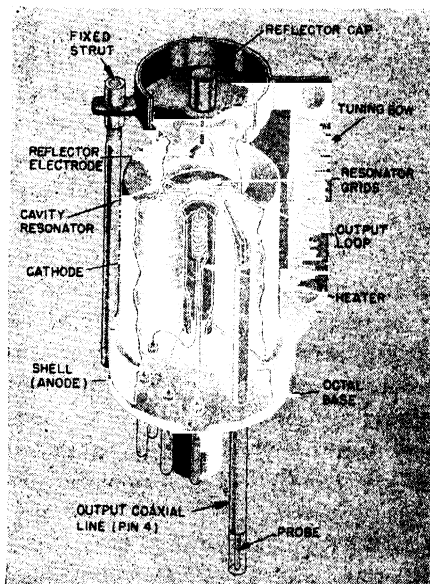



FIG. 437*b*.—Internal construction of type 723 reflex klystron.

modulated) oscillators used in r-f spectrum analysis. Stability of the output frequency depends on accurately stabilized power supply.

Among the reflex klystrons used for test oscillators are types 417A, 707A, and 723A, all of which find use as local oscillators in receivers. They provide power output levels in the hundreds of milliwatts. For higher power output, double-cavity klystrons may be used, such as type 410R (3K30), which has a maximum power output of about 40 watts at 10-cm wavelength. Figures 437, 438, and 439 show typical klystron assemblies.

Modulation of test oscillators is often required. Amplitude

modulation at audio frequencies is necessary when amplification of the test signal is required to achieve high sensitivity. Such audio amplitude modulation generally should not be accompanied by frequency modulation. To produce amplitude variations free of frequency variations, it is customary to apply a rectangular wave to the repeller electrode. The oscillator is thereby forced completely out of oscillation during one-half of the audio cycle, and oscillates steadily at one frequency during the other half. Frequency modulation, substantially free of amplitude variations, is required for spectrum analysis. The application of a low-amplitude sawtooth wave to the repeller electrode serves this purpose.

FIG. 438a.—Type 3K30 (410R) double-resonator klystron, tunable over range from 2700 to 3330 megacycles, with a power output (c-w) of from 20 to 40 watts.

Test oscillations at lower frequencies (500 to 2,500 megacycles) may be generated conveniently in disk-seal ("light-house") tubes and appropriate circuits. One such circuit is the "coaxial butterfly," a coaxial resonant cavity whose outer wall is composed of two adjustable segments. Such a test oscillator assembly is shown in Fig. 440. Audio modulation of such oscillators may be accomplished by conventional methods, such as plate or grid modulation.

215. Power-measuring Devices.—Power measurements at low levels (from 1 to 1,000 μwatts) are generally carried out by bolometers or calibrated crystal rectifiers. The bolometer comprises a circuit element whose ohmic resistance changes when it absorbs heat from the r-f source. The change in resistance is measured, usually in a resistance bridge, and the r-f energy computed from a resistance vs. power curve.

One common form of bolometer is the "Littelfuse," an instrument-protecting fuse having a fine platinum wire as the fuse element. For sensitive power measurements fuses rated at 5 or 10 ma are used. The slope of the resistance-power curve of this fine (0.00015 in. diameter) wire is 2,000 to 5,000 ohms

per watt. Consequently to measure a power level of a few microwatts, it is necessary to measure resistance changes of a few milliohms, which can be done readily on a balanced resistance bridge. Unbalanced bridges, which give substantially linear indication of the power, are also used

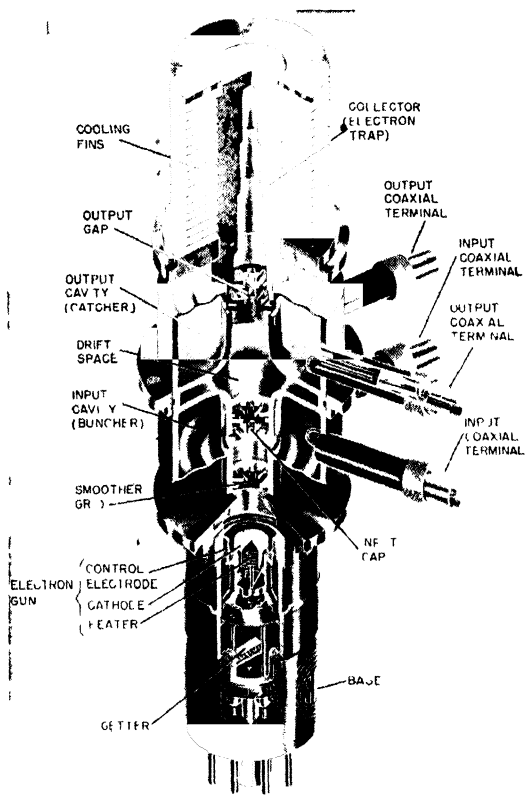



Fig 438*b*—Internal structure of 3K30 klystron

Another form of bolometer is a mixture of metallic oxides known as a "thermistor" This material has a negative resistance-power slope about 10 times as great as the platinum wire (20,000 or more ohms per watt), it is more rugged than the fuse, and it is less subject to damage by overload The thermistor is used in the form of a small bead supported on wires and mounted in a cartridge. A typical unbalanced thermistor bridge is shown in Fig 441 Using a 0-200-μamp meter as the indicator,

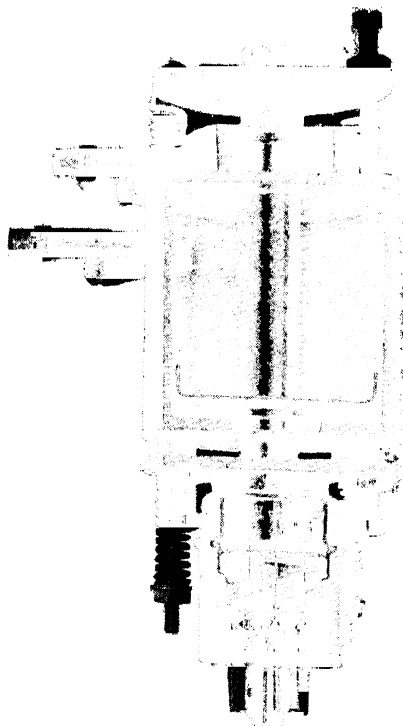



FIG 439 —Cutaway view of type 2K46 frequency multiplier klystron Input 3000 mc, output 9000 mc, tunable over a 10 per cent range

FIG 440 —Coaxial butterfly oscillator element and disk-seal tube.

this bridge indicates a maximum r-f power of 1 mw, and it has a linear calibration.

Since the thermistor is sensitive to external temperature changes, as well as to those produced by the absorbed r-f energy, it is necessary to compensate for the ambient temperature. In the bridge shown, d-c and a-c voltages are applied to the

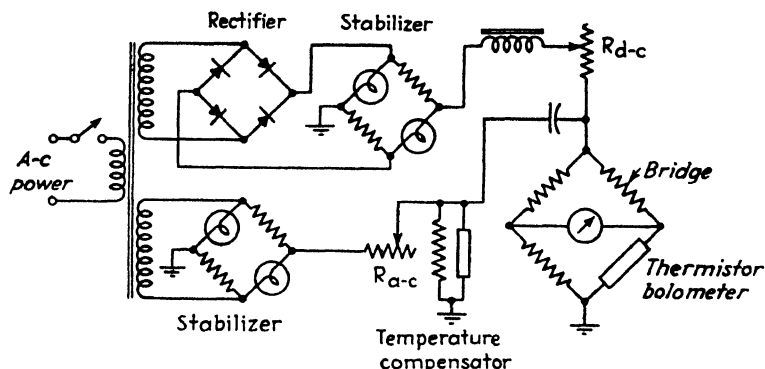



FIG. 441.—A-c operated thermistor bridge for measuring low-power r-f sources.

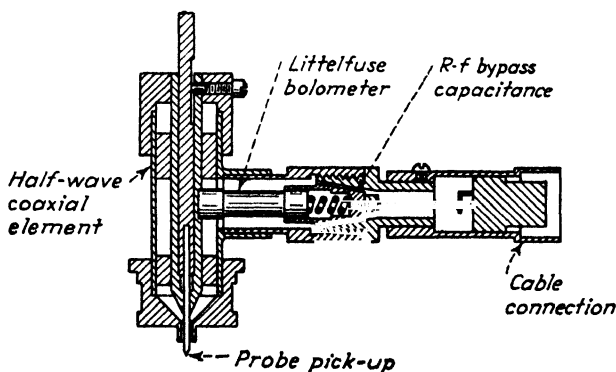



FIG. 442.—Fuse bolometer probe (fixed tuned, 10-cm).

bridge, both being regulated by passage through auxiliary tungsten-filament lamp bridges. The detector bridge is first brought to balance with the alternating voltage removed, by adjustment of the series resistance R_{d-c} until the direct current applied raises the temperature of the bridge elements above ambient temperature. Then the a-c voltage is applied and the microammeter indicates the r-f input at a sensitivity of about 5 μw per μa deflection.

The Littelfuse or thermistor bolometer is mounted in a coaxial or waveguide mount so as to form a matched load. Typical structures are shown in Figs. 442 and 443. Broadband mounts have been designed that avoid the need of tuning over a 2:1 frequency range.

The calibrated crystal rectifier is not so sensitive as the bolometers just considered, but it has the advantage of simplicity. The crystal is the standard silicon cartridge used in microwave

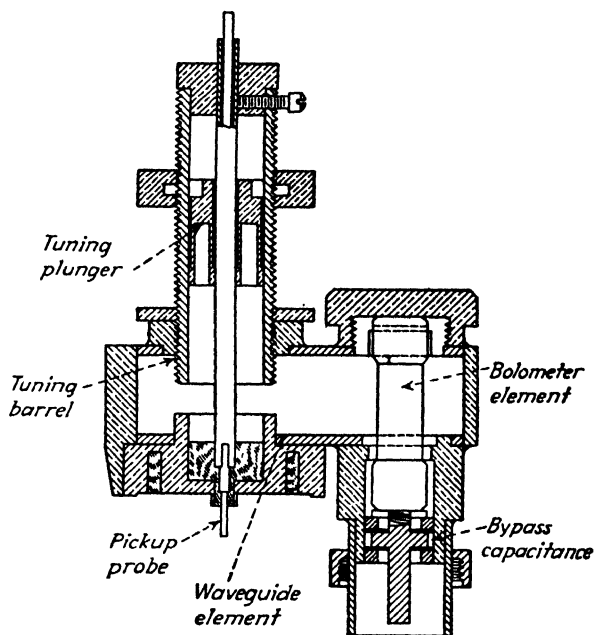



FIG. 443.—Fuse bolometer probe, variable tuned 3-cm type.

mixers and is connected to a microammeter of about 0 to 100 μa range. The crystal mount includes a thin dielectric section that by-passes the r-f components. The assembly is calibrated against a bolometer in a balanced bridge. A typical calibration curve is shown in Fig. 444. One of the disadvantages of the calibrated crystal rectifier is its susceptibility to gross shifts in calibration due to mechanical shock or electrical overload. This is particularly possible when measuring pulsed power, since the peak power may do damage even when the average power level is low. For this reason crystals should be used only when they can be subjected to frequent calibration checks.

For high-power levels (average power in the tens or hundreds of watts), two useful power-measuring devices are the water calorimeter and the Johnson power meter. The calorimeter is illustrated in Fig. 445. It consists of a water chamber, with water

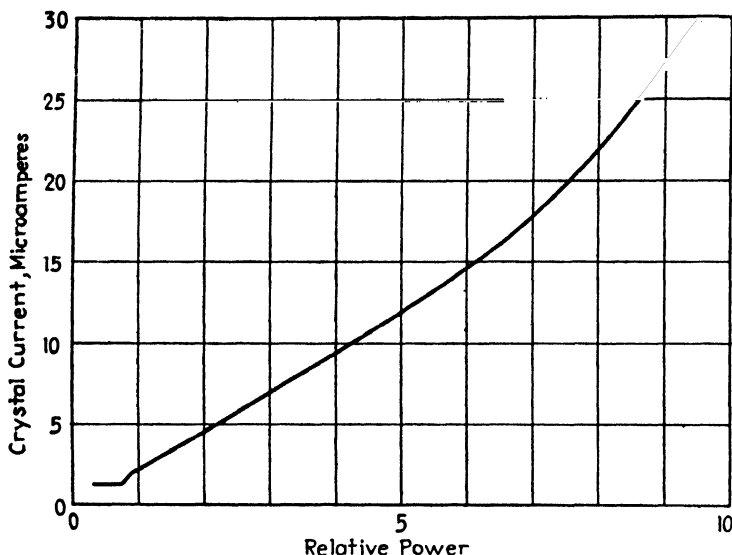



FIG. 444.—Typical silicon crystal calibration.

inlet and outlet, separated from the r-f power source by an insulating segment. A quarter-wave thickness of the insulating segment forms a match between the empty coaxial line or waveguide on one side and the water-filled element on the other.

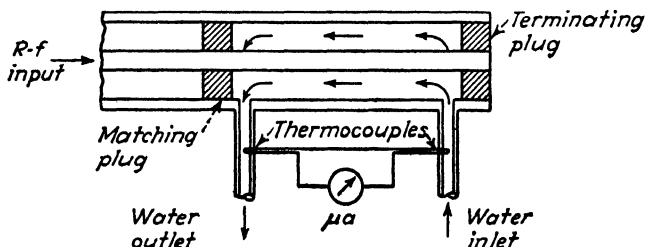



FIG. 445.—Water calorimeter for measuring large amounts of r-f power.

The normalized impedance of the coaxial line or waveguide, when filled with the insulating material, must be the geometric mean of the impedances when empty and when water filled. Glass or mycalex serves the purpose.

The calorimeter measures the power at a constant known rate of water flow through the water chamber, by the change of temperature between water input and output. The power p absorbed in the water, in watts, is equal to the product of the mass m of water, in grams, passing through the chamber per second, the heat capacity of the water (4.19 watt-sec per g per deg centigrade), and the change in temperature ΔT in degrees centigrade. Thus

$$p = 4.18m\,\Delta T \qquad \text{watts} \tag{505}$$

Since the specific gravity of water is one gram per cubic centi-

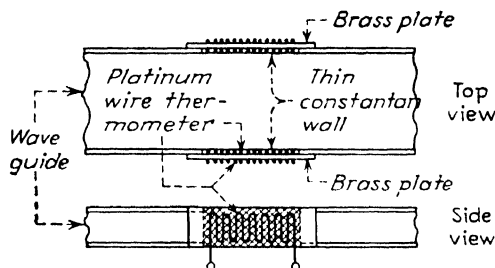



FIG. 446.—Johnson power meter.

meter, this relation may also be written

$$p = 4.18v\,\Delta T \tag{506}$$

where v is the volume of water flow in cubic centimeters per second.

The difference in temperature is measured by two thermocouples, in the inlet and outlet, connected in series with a galvanometer, calibrated in temperature difference in degree centigrade. The rate of water flow is readily found by collecting the outlet water in a graduate, over a measured length of time.

This method of measuring power is absolute and may be refined to give accuracies better than 1 per cent. Any amount of power up to several thousand watts, average, may be dissipated in the water load by suitable choice of the rate of flow.

When it is desired to measure large amounts of power, without consuming all of the power measured, a sampling technique may be employed. The Johnson power meter, shown in Fig. 446, is an example. It is applicable to rectangular waveguide systems. A section of the narrow face of the guide is removed

and replaced by a thin (0.015-in.) section of constantan (a copper-nickel alloy). The high resistivity of the constantan, relative to that of the copper guide wall, causes it to assume a slightly higher temperature. Since the section is thin, it assumes the temperature difference rapidly, that is, in less than a minute after application of the power. The temperature difference is measured by a platinum-wire thermometer. One 200-ohm coil of fine platinum wire is placed in thermal contact with the constantan, while another similar coil is placed on the copper guide wall, near by. The two coils form the arms of a Wheatstone bridge, operated in unbalanced condition. The unbalance current is indicated in a 0 to 200 microammeter. Calibration is effected by comparison with a water calorimeter. This power meter requires no water supply, is accurate to about 5 per cent, and consumes less than 1 per cent of the power transmitted. For the latter reason it serves well as a monitor of power flow, since it can be permanently fixed to a waveguide transmission system without wasting any substantial amount of power. High-power levels may also be measured by the low-power devices previously described, provided that suitable calibrated attenuators are available to reduce the power from the level of watts to milliwatts or microwatts. Attenuators for this and other purposes are described in the next section.

216. Attenuators and Fixed Loads.—Ultrahigh-frequency and superhigh-frequency attenuators serve to introduce known amounts of attenuation in power measurements, or to isolate one piece of equipment from another. The latter function is performed by the attenuator "pad." Pads are usually inserted (1) after signal generators to prevent the test circuit from changing ("pulling") the oscillator frequency, or (2) preceding power indicators to avoid reaction of the indicator on the standing wave pattern.

Among the variable attenuators is the "waveguide-below-cutoff" type. Cutoff attenuators consist of a segment of waveguide whose cross-sectional dimensions are smaller than the cutoff values. Power is fed to one end of the segment by a coupling loop or plate, and removed from the other end by a similar connection. If the propagation within the segment is limited to one mode, it is possible to compute accurately the attenuation suffered by the signal in passing through the segment.

The waveguide segment is ordinarily of circular cross section, operated either in the $TE_{1,1}$ mode or $TM_{0,1}$ mode. The corresponding attenuations, below cutoff, are

$$\text{For } TE_{1,1} \qquad \alpha = \frac{32}{d}\sqrt{1 - \left(\frac{1.7d}{\lambda}\right)^2} \text{ db/cm} \tag{507}$$

$$\text{For } TM_{0,1} \qquad \alpha = \frac{42}{d}\sqrt{1 - \left(\frac{1.3d}{\lambda}\right)^2} \text{ db/cm} \tag{508}$$

where d and λ are the diameter of the guide and the wavelength, both in centimeters. The $TE_{1,1}$ mode is excited by a coupling loop, the $TM_{0,1}$ by a flat plate. Figure 447 shows the two forms in a coaxial line.

FIG. 447.—Waveguide below-cutoff attenuators.

While the cutoff attenuator is simple and capable of self-calibration, it suffers from two drawbacks. First, the waveguide segment does not dissipate power, and therefore reflections are set up at the input termination. The reflected waves must be absorbed in some other element, such as dissipative matching sections (lengths of "lossy" cable), etc. The second disadvantage is that the coupling loops or plates must not be brought too close together, and the minimum attenuation (insertion loss) of the attenuator is correspondingly high. If the loops or plates are in close proximity, coupling may occur between them at higher modes, which affects the calibration in a complicated, nonlinear manner. In practice the minimum insertion loss of cutoff attenuators is about 20 db, and the linear portion of the calibration begins at about 30 db. Above this value the attenuation varies linearly with the distance between loops or plates. Very large amounts of attenuation may be introduced by employing long waveguide segments. Figure 448 shows methods of introducing cutoff attenuation between waveguide elements, using coaxial lines and probes as connecting links.

The other class of attenuator comprises the dissipative types that consume part of the energy passed through them. Dissipa-

tive attenuators are constructed in various ways, depending on the shape and form of the transmission system. In coaxial systems, dissipative elements are commonly formed on a glass rod that has the same diameter as the inner conductor of the line. The glass is coated with a very thin coating of platinum or carbon, which is much thinner than the skin depth ($=10^{-4}$ cm at 3,000 megacycles in copper). Since the coating is resistive, heat is dissipated as the energy passes. The power level of this type of attenuator is limited by the ability of the glass member to conduct heat. Generally no more than a few watts can be

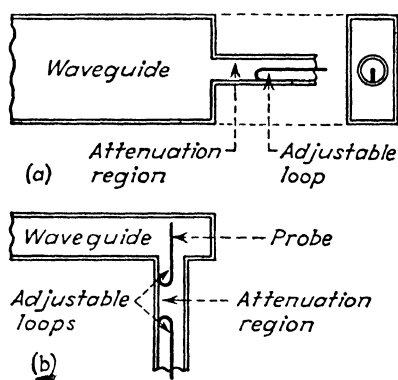



FIG. 448.—Application of below-cutoff attenuators to waveguides

handled, but the attenuation may be as high as 30 db in a 10-in. section.

It should be remembered that dissipative elements in transmission lines introduce a complex characteristic impedance (cf. Eq. (151*a*), Sec. 65) that varies with the frequency. It is advantageous to match this complex impedance to the real characteristic impedance of the nondissipative line. To accomplish this, the glass may be coated with two thicknesses of material, a thin coat for the dissipative element coupled at either end by somewhat thicker coats, which serve to match the dissipative element to the transmission system.

In the simple form just described the platinized or carbonized attenuators have fixed values. They may be constructed in variable form by the use of a metal sheath which slides over the dissipative element, or by a metal rod which slides inside the glass element. Such variable attenuators must be calibrated by reference to attenuator standards, usually of the cutoff type.

A material capable of introducing a very high degree of attenuation in a short space is finely divided iron powder, known as "polyiron." This material has high permeability at radio frequencies up to about 30 megacycles and very high attenuation for r-f energy at frequencies higher than 1,000 megacycles. In the powdered form this material will attenuate at the rate of 5 to 10 db per cm at 3,000 megacycles. In the molded form, values run as high as 30 db per cm at 3,000 megacycles and correspondingly higher at 10,000 megacycles. One of the important uses of molded polyiron is as a choking element. The material may be placed around the inner conductor of coaxial input leads in test equipment where signal leakage must be kept to a low value. Rubber dielectric is also capable of introducing high degrees of attenuation at microwave frequencies.

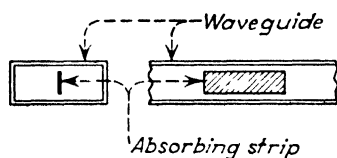



FIG. 449.—Strip attenuator.

Dissipative attenuators for waveguide systems generally make use of a resistor element in strip form, such as a layer of bakelite plastic coated with graphite. In rectangular guides, the strip is placed along the axis of the guide and parallel to the narrow wall as shown in Fig. 449. Maximum attenuation occurs (in the $TE_{1,0}$ mode) when the strip is in the center of the guide, since the electric field is greatest along this plane. The attenuation increases with decreasing resistivity of the strip coating and with increasing length of strip.

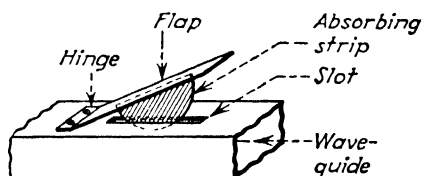



FIG. 450.—Flap-type adjustable attenuator.

A variable strip attenuator uses a "flap" carrying a strip that protrudes into the guide through a slot (Fig. 450). Adjustment of the flap position varies the attenuation, which must be calibrated against a standard. The attenuation may also be varied by moving the position of the strip laterally in the guide from the center plane (maximum electric field) to the side walls

(zero field). A screw mechanism for this purpose is illustrated in Fig. 451.

The power level of the waveguide strip attenuator, like the platinized coaxial form, is limited to a few watts by poor heat conduction. Larger amounts of power may be dissipated by solid dissipative elements that fill the whole cross section of the guide and extend over several wavelengths. These "plug" attenuators (Fig. 452) generally are tapered at either end to

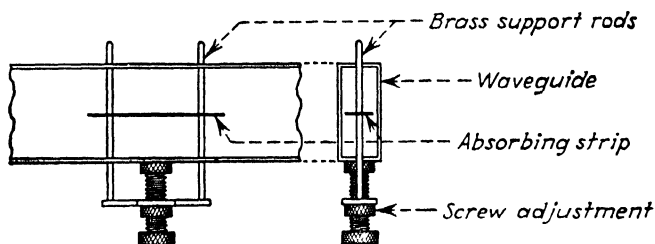



FIG. 451.—Adjustable strip attenuator.

improve the impedance match with the remainder of the system. The plugs may be made of bakelite, of a graphite suspension in plaster, or even of wood.

Closely allied to calibrated attenuators are dissipative elements used to terminate a high-power test system. These are generally known as "loads." The water calorimeter previously described is an excellent load, but the running water system is inconvenient. "Dry" loads are formed by filling the coaxial line or waveguide with sand coated with colloidal graphite (aquadag). In the coaxial form (Fig. 453) the material is

FIG. 452.—Plug attenuator in circular waveguide.

prevented from flowing into the pipe by a layer of cement, and the impedance is matched, approximately, by the smaller diameter of the inner conductor in the sand-loaded segment. In waveguides, a similar sand mixture is used and the impedances are matched by tapering the sand surface over at least two wavelengths (Fig. 454). In sand loads most of the power is absorbed close to the surface on which the power impinges. To assist in dissipating the heat, cooling fins are generally soldered

on the line or guide, in the region of greatest power concentration. Air cooling is sometimes used in high-power systems. Peak power greater than about 50 kw is not recommended in coaxial sand loads since the cement seal may arc over. High peak and average powers may be dissipated readily in the waveguide form.

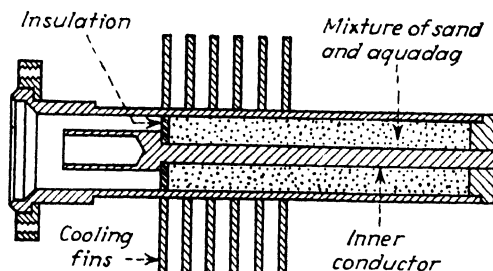



FIG. 453.—High-power sand load.

Another form of load which can dissipate a large amount of power is the antenna load, which may take several forms. An ordinary dipole may be used, but is subject to changes in impedance due to reflecting objects in the vicinity. A similar load is the so-called "space-matcher," which matches the open end of a waveguide to the adjacent empty space by an inductive window placed near the opening. This type of load also suffers from the uncertain effect of near-by reflections.

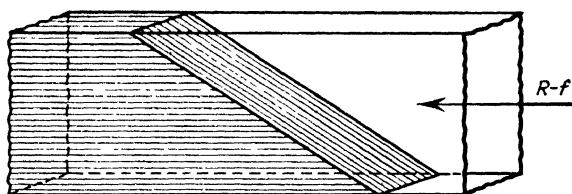



FIG. 454.—Tapered plug load in rectangular guide.

An antenna load that eliminates reflections is the "black-box" load. This is a broadband dipole radiator in a rectangular box. The inner surface of the box is composed of fabric coated with graphite, the lateral (surface) resistivity of which is 377 ohms per unit square (the characteristic impedance of empty space). One-quarter wavelength behind the fabric is a metallic wall (copper). This combination, known as a "Salisbury screen," absorbs energy of the wavelength for which it is designed, hence

the name "black box." The energy radiated by the dipole is wholly absorbed in the graphite surface and no reflections occur. Since the black-box load is large and cumbersome, and since it operates over a narrow frequency band, it is not much used.

217. Frequency Standards and Wavemeters.—Calibration of frequency in the uhf and shf spectra is referred to primary standards that derive their calibration directly from standard time. A block diagram of such a primary frequency standard is shown in Fig. 455. The basic timing source is a 100 kc quartz crystal, which is carefully mounted and temperature controlled. The fiftieth harmonic of this crystal is compared with the 5-mc

FIG. 455.—Block diagram of microwave-frequency calibration system.

signals of station WWV. An auxiliary oscillator is then adjusted to the seventieth harmonic of the crystal and its output multiplied to the microwave region by three tripler and four doubler stages.

Such primary standards are used only for laboratory calibration purposes. Portable wavemeters serve as suitable secondary standards. The two common types of wavemeters are the coaxial and the cylindrical cavity. The coaxial variety is self-calibrating but generally displays lower Q than the cylindrical and thus has lower frequency resolution.

As in lower frequency practice, wavemeters may be used either as transmission or absorption instruments. The transmission type has two terminals and transmits a maximum signal at resonance, whereas the absorption type has one terminal and reflects a high impedance at resonance. This high impedance

may be transformed by a quarter-wave section to low impedance. A sensitive power-measuring device, such as a crystal probe and microammeter, is used to indicate the resonance point. Typical wavemeter connections are shown in Fig. 456.

Wavemeters are coupled to the external circuit by a coupling loop, placed at a point of maximum magnetic field. The plane

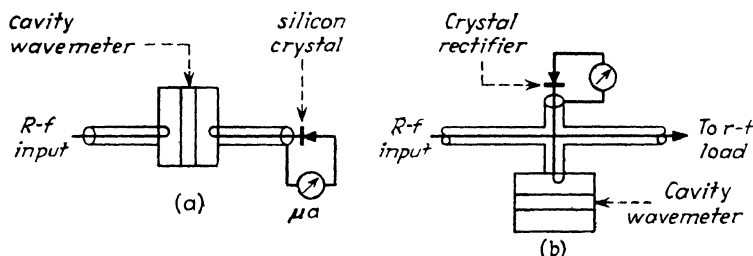



Fig. 456.— Methods of indicating wavemeter resonance.

of the loop is rotatable in some types to permit adjusting the degree of coupling. The principal mechanical element of a wavemeter is a sturdy micrometer movement for adjusting the length of the coaxial line or the volume of the resonant cavity. Standard micrometer movements are used.

For most purposes, in the microwave region, it is considered sufficient to measure wavelength accurate to about plus or minus

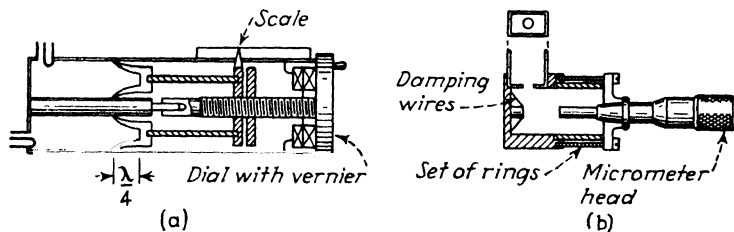



Fig. 457.—Coaxial wavemeters.

0.001 cm, or about 0.02 per cent at 10 cms wavelength. To meet this requirement the micrometer movement must be capable of indicating changes in distance of this order (0.001 cm), and the Q of the wavemeter must be of the order of 5,000 (response curve down 3 db at 0.02 per cent off resonance). Coaxial wavemeters meet this limit and well-designed cylindrical cavity meters exceed it by a factor of 8 to 10.

Typical constructions of coaxial and cylindrical wavemeters

are shown in Figs. 457 and 458. The cavity types are complicated by the fact that the resonant mode used must not involve current lines crossing the wiping surface at the edge of the movable wall. High modes are thus generally required, and multiple mode excitation is possible. Auxiliary absorbing structures, in the form of wire grids or absorbing layers, are used

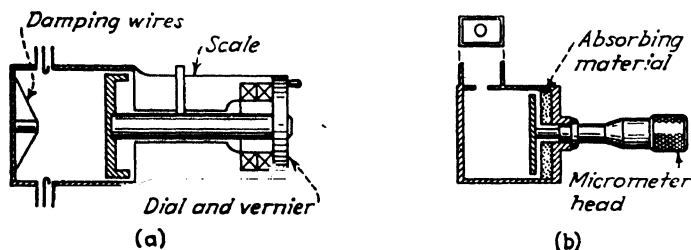



FIG. 458.—Cavity wavemeters.

to damp out the undesired modes. Fixed cavities, calibrated by reference to a primary standard, are often used as secondary standards in the laboratory.

A device closely resembling the cavity wavemeter, but used for a different purpose, is the echo box. This device gives an over-all indication of radar system performance. It is essentially a resonant cavity connected in some way to the radar transmitter and shock excited by each transmitted pulse. At the conclusion of the pulse, the excitation continues as a damped continuous wave that dies away exponentially, persisting at a perceptible level for a number of microseconds. The echo-box signal is picked up by the radar receiver and indicated on the type A indicator as an exponential curve. If the power level of the echo-box excitation is known, the system sensitivity may be computed by the duration of the exponential response. The normal duration of the response, known as the "ringing time" of the echo box, is the interval from the deionization of the t-r switch (immediately after the transmitted pulse is completed) to the time at which the exponential signal becomes lost in the receiver

FIG. 459.—Echo-box trace on type A indicator.

noise. Typical ringing times are from 10 to over 100 μsec, depending on the Q of the echo box and related system constants, such as transmitter power, antenna gain, and receiver noise figure. A typical echo-box signal, displayed on an A scope, is shown in Fig. 459.

The two forms of echo box are the tuned and untuned types. The tuned box (Fig. 460*b*) is very similar to a cylindrical wavemeter, operated in the $TE_{0,1,1}$ mode. In the interest of a long ringing time, the cavity is as large as is possible, consistent with suppression of higher modes, and its interior may be silver-plated to lower the internal losses. The box may be coupled to the radar by a dipole pickup antenna placed near the radar

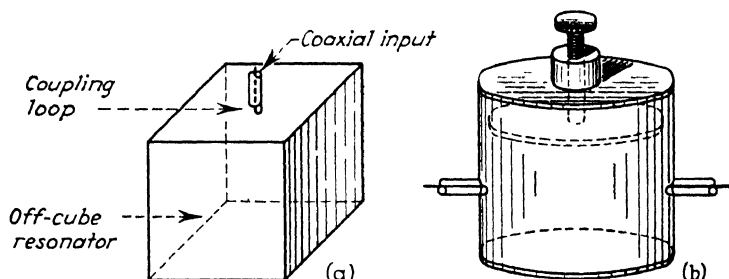



FIG. 460.—Untuned rectangular (*a*) and tuned cylindrical (*b*) echo boxes.

radiator. The position of the pickup must be standardized if comparable results are to be obtained at different times. Direct connection with the r-f transmission system of the radar may also be made through a directional coupler to avoid the effects of space reflections and standing waves.

The untuned box (Fig. 460*a*) is one made large enough to insure excitation of at least one, and possibly several, modes within the tuning ranges of the radar transmitters and receivers to be tested. No attempt is made to suppress the higher modes in this type. The shape is generally that of a cube, but the dimensions are made slightly irregular, since a perfect cube would allow excitation of many modes at a single value of wavelength. To insure a copious number of different resonant wavelengths, the linear dimensions of the box are usually chosen to be 10 wavelengths long, corresponding to l, m, and n subscripts of that order. This requirement gives rise to the principal disadvantage of the untuned echo box, that is, its awkward size.

The large size is advantageous, however, since it gives rise to a very high value of Q and correspondingly long ringing time. A copper cubical box about 3 ft on a side, operated at 3,000 megacycles, has an unloaded Q of about 500,000 and a ringing time in excess of 100 μsec on a 50-kw radar system.

The logarithmic decrement of the echo-box signal implies a proportional relationship between system performance expressed in decibels and observed ringing time. An increase (or decrease) of N decibels in transmitter power, antenna power gain, or receiver power sensitivity produces an increase (or decrease) in the ringing time by an amount Δt, in accordance with

$$\Delta t = \frac{0.23NQ}{2\pi f} \qquad \mu\text{sec} \tag{509}$$

Here Q is the loaded Q value of the echo box and f is the carrier frequency in megacycles. Thus at 3,000 megacycles, the signal of an untuned echo box of loaded $Q = 100{,}000$ will ring for one additional microsecond for every decibel improvement in system performance.

Echo boxes need not be connected to the radar receiver if there is any doubt as to the receiver performance. A calibrated silicon crystal detector and microammeter as indicator will indicate the echo-box response, averaged over several pulses, and thus indicate the transmitter performance. If the echo box is of the tuned type, the crystal detector may be used to explore the spectrum of the transmitter output by tuning the box through the transmitter band and noting the variation in the crystal current.

218. Standing-wave Detectors.—The measurement of standing waves in coaxial and waveguide transmission systems is essential to assure optimum performance of a radar system. As indicated in Chap. III (Sec. 67 to 71 inclusive) standing waves must be minimized for maximum power transfer and to avoid excessive peak voltages in the transmission system. Moreover the uhf and shf oscillators employed in radar, particularly cavity magnetrons, operate with high efficiency and stability only into a load of specified impedance. The matching of the oscillator to this preferred load is carried out by standing-wave measurements.

Standing waves are measured by inserting a short probe into the coaxial line or waveguide and moving the probe along the

line or guide. A detector connected to the probe indicates the intensity of the electric field. When standing waves are present, regions of maximum and minimum intensity are observed, separated by one-quarter of the wavelength within the transmission system. The ratio of the maximum field intensity to the minimum field intensity is the standing-wave voltage ratio. The position of the maximum or minimum, relative to a reference in the system (for example, load or generator termination), indicates the phase of the standing-wave pattern. The standing-wave ratio and the nodal position in electrical degrees may be plotted on a circle diagram to indicate the impedance or admittance level at any point in the transmission system.

To permit the moving of the exploring probe along the line, it is necessary to provide a longitudinal slot in the outer wall of the coaxial line or guide, and this slot must be so positioned

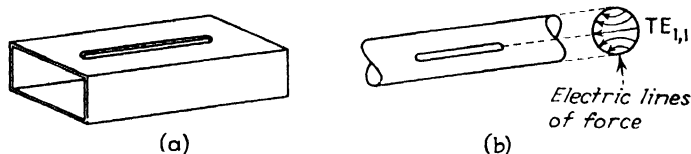



FIG. 461.—Slotted sections in rectangular and circular waveguide.

that it does not interfere with the normal propagation of energy. Segments of coaxial line or waveguide provided with such slots are known as "slotted sections." The length of the slot must be at least one-half wavelength to permit reaching two adjacent maxima; usually the length is several wavelengths. In the coaxial line, operating in the *TEM* mode, a slot parallel to the axis of the line lies parallel to the lines of current, and hence does not affect propagation. Likewise in rectangular waveguide, operating in the $TE_{1,0}$ mode, the slot may be placed in the center of the wide face parallel to the axis, without disturbing the current lines. In circular waveguide, a longitudinal slot may be used in the $TM_{0,1}$ and $TE_{1,1}$ modes. In the latter mode, however, the slot must be placed in a particular position, shown in Fig. 461*b*, relative to the polarization of the electric vector.

Extending into the slot is the probe, the length of which must be short (generally not more than one-twentieth wavelength) to avoid introducing reflections on its own account. The probe connects directly with a power-measuring device, such as a

silicon crystal or one of the bolometers previously described (Littelfuse or thermistor). The power-measuring device develops a d-c or a-f (audio-frequency) signal that is some known function of the electric intensity at the probe. This d-c or audio output is conveyed through a length of flexible coaxial cable to an indicator, such as a microammeter, or audio amplifier, rectifier, and meter. The meter may be calibrated directly in electric

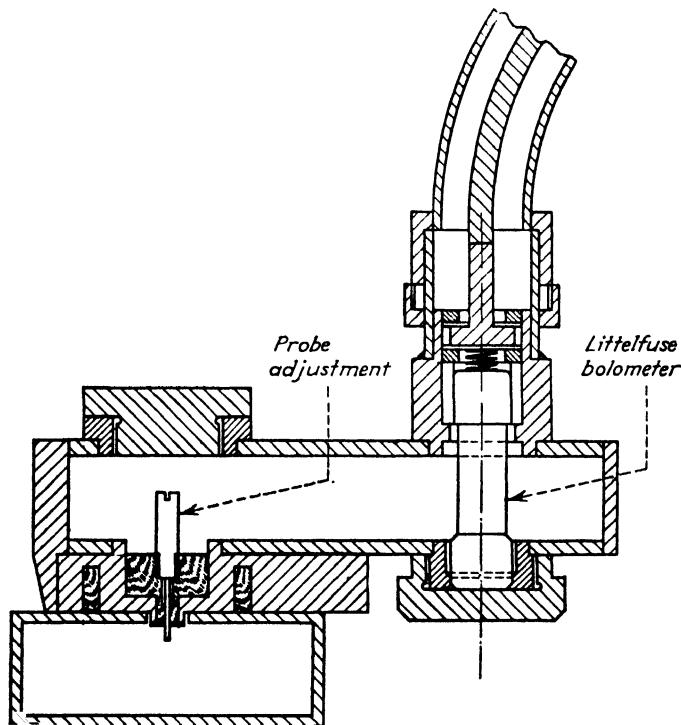



FIG. 462a.—Standing-wave detector for rectangular guide.

intensity ("voltage") units to facilitate forming the standing wave voltage ratio. Figure 462 shows typical constructions.

Since the ratio of electric field intensities is measured indirectly in terms of the power level indicated by the detector, it is essential to know the relationship between voltage input to the detector and the d-c or a-f output indication. The silicon crystal detector is very closely a square-law device, that is, the output direct current is proportional to the square of the electric intensity at the probe. The standing-wave voltage ratio is then simply the

square root of the ratio of maximum and minimum currents. The square law applies quite accurately when the crystal current lies between 10 and 50 μamp, which permits measuring standing-wave voltage ratios up to $\sqrt{50/10} = 2.23$. Since accurate measurement of standing-wave ratios higher than 2 are seldom required, the crystal can generally be assumed to follow the square law, provided the maximum and minimum currents do not fall outside the above limits. The voltage across the probe can be adjusted to fall within the desired limits by varying the output of the test oscillator or the depth of penetration of the probe into the coaxial line or guide.

FIG. 462b.—Slotted-section standing-wave detector for coaxial lines.

If the calibration curve of the crystal is in doubt, it may be checked by operating it in a transmission system containing a calibrated attenuator. Another simple calibration procedure is to short-circuit the slotted section at its output end. The probe picks up zero field at the short circuit, and the field increases with distance from the short circuit, varying as $\cos 2\pi l/\lambda$. Knowing the wavelength λ, it is possible to compute the intensities at different lengths l from the short circuit, and to plot these intensities, on log-log coordinates, against the corresponding observed d-c outputs. The slope of the resulting curve gives the exponent of the law followed by the crystal; usually the slope is close to 2, representing the square law.

When standing waves must be measured in transmission systems at very low power levels, the crystal and microammeter combination is not sufficiently sensitive, and therefore it is necessary to employ amplification. Since d-c amplification is difficult to stabilize, it is customary to modulate the test oscillator (by a rectangular wave in the case of a reflex klystron, as previously described) at audio frequencies and to amplify the crystal output in a stable (inverse-feedback) audio amplifier.

The output of the audio amplifier is rectified in a conventional diode and applied to a milliammeter. The law followed by the crystal at low levels may not be that of a simple exponent, so it is customary to calibrate the crystal, amplifier, and indicator system by reference to a calibrated variable attenuator.

When a bolometer is used as the detector, the change in resistance developed by the absorbed power may be measured in a bridge, as in the previously described absolute power-measuring systems. But since the standing-wave ratio is a relative quantity, absolute values of power are not required, and the bridge indicator is usually dispensed with as being too cumbersome for the purpose. In its place, the test oscillator is modulated at an audio frequency, and the resistance of the bolometer element is thereby varied at the a-f rate. The change in resistance is indicated by passing a small direct current through the bolometer element. This current, modulated at the audio frequency, is passed by a coupling transformer to a stable audio amplifier.

The audio amplifier used should have a linear relationship between input and output, should give full scale output with a voltage input of the order of 1 to 10 μv, depending on the sensitivity of the crystal or bolometer detector, and should have low internal noise and hum. To minimize false indications due to hum, it is customary to limit the l-f gain of the amplifier. In any event, the indication of the output meter with zero input should be noted and subtracted from the values indicated when the excitation is present.

219. Tuners, Couplers, and Power Dividers.—Standing-wave compensators, known as "tuners," consist of reactive elements of adjustable magnitude and phase that may be introduced in a transmission system to produce a standing wave equal and opposite to the one to be removed. Brief mention of such reactive elements has been made in Chap. III, Secs. 69 and 81. Here we consider their physical forms and methods of adjustment.

One of the most widely used tuners is the stub tuner, consisting of one or more adjustable side stubs that add susceptance to the coaxial line or waveguide. The added susceptance is adjusted to be equal and opposite to the susceptance causing the standing wave, when both are transformed to the same point in the line. A single stub can be used for this purpose only when its position

along the line is adjustable and this is inconvenient from a mechanical point of view. To avoid shifting the position of the stub, two or more stubs may be used, spaced at some fixed interval. The double-stub tuner is the commonest form.

The double-stub tuner consists of two tunable stubs spaced three-eighths of a wavelength (the wavelength in guide, in the case of the waveguide tuner). The susceptance presented by one stub is transformed, on the admittance circle diagram, to the position of the other by rotating the admittance point three-quarters of a turn counterclockwise about the center of the circle diagram. To understand the action of the double-stub tuner, consider the admittance circle diagram (of the Smith type, see

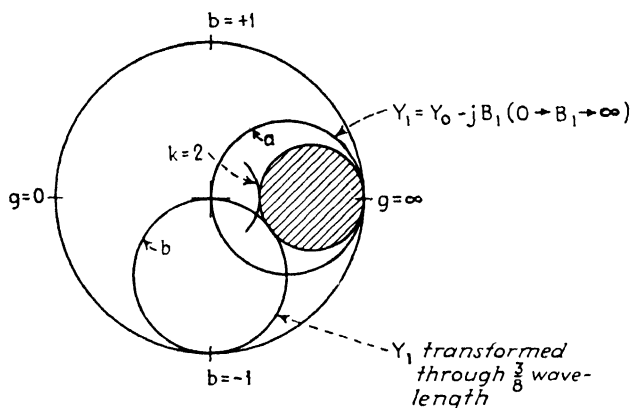



FIG. 463.—Circle diagram of double-stub tuner.

Sec. 70) in Fig. 463. It is desired to match a load of arbitrary terminating admittance to the characteristic admittance Y_0 of the line. At the input to the tuner (at stub 1), we add the susceptance of the stub B_1 to the characteristic admittance. As the position of the plunger is varied through a quarter wavelength, B_1 varies from zero to infinity and the net admittance at the first stub revolves about the admittance circle corresponding to $Y_1 = Y_0 - jB_1$, as shown at *a* in the diagram. This circle lies tangent to the *b* axis and the outer limit of the circle diagram. The admittance values on this Y_1 circle are transformed to the position of the second stub, three-eighths wavelength away, by rotating the Y_1 circle three-fourths turn in a counterclockwise direction to the position shown at *b* in the diagram. Now by adjustment of the susceptance of the second

stub B_2 we can transform any value on the Y_1 admittance circle to some other point of the circle diagram, thus matching the arbitrary load admittance. The effect of adjusting B_2 is to translate any point on the rotated Y_1 admittance circle, along a line of constant conductance, to any other point. Examination of the diagram (Fig. 107) shows that such translation can reach any point of the circle diagram, except the points within the shaded circle, which is (1) bounded by the value $g = 2$ and (2) tangent to the rotated Y_1 admittance circle. The values of admittance within the $k = 2$ circle can, therefore, be matched under any circumstance. In other words, if the standing-wave

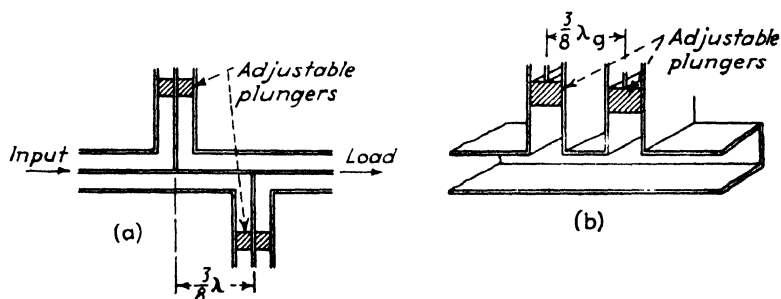



FIG. 464.—Coaxial and waveguide double-stub tuners.

voltage ratio k is 2 or less, as is generally the case, the standing wave can always be removed by a double-stub tuner of this type. If the standing wave is larger than this, it can be removed if the position of the nodal point is such as to place the voltage minimum outside the shaded area.

The physical form of double-stub tuners is shown in Fig. 464. In the coaxial form, the stub susceptance appears in shunt, and the admittance diagram is used in analyzing its action. In rectangular waveguides, however, two choices are possible. The stubs may be placed in the wide surface of the guide. In this case they interrupt the current flow, that is, the current must travel up into and out of each stub. These stubs represent a series reactance, and are treated by the *impedance* form of diagram. If the stubs are inserted in the narrow face of the guide, the voltage gradient between walls is common to waveguide and stubs. This arrangement introduces shunt reactance, which is treated with the *admittance* diagram.

Another type of tuner, used widely in coaxial as well as wave-

guide systems, is the slug tuner. Two slugs, made either of dielectric material or of metal, each one-quarter wavelength long, are placed within the coaxial line or waveguide. The slugs are mounted so that the separation between them, as well as their position along the line, can be adjusted. By proper positioning of the slugs and their separation, it is possible to compensate large standing waves (standing-wave voltage ratios of 5 to 10).

To understand the action of the slug tuner, consider two quarter-wave dielectric slugs in a coaxial line, as illustrated in Fig. 465. The dielectric constant of the material is K. Consequently the characteristic impedance of the line section con-

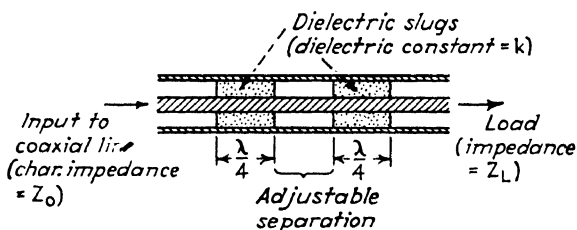



FIG. 465.—Dielectric slug tuner.

taining the slug is $Z_0/\sqrt{K}$. A quarter-wave segment of such a line transforms the line impedance Z to another impedance, according to Eq. (166), equal to $(Z_0^2/K)/Z_0 = Z_0/K$. The presence of such an impedance in the line produces a standing-wave voltage ratio of $k = K$. Thus if a standing wave of this ratio already exists, the slug can be so positioned as to compensate for the existing wave.

Such compensation is of limited interest, since it is effective only at one value of standing-wave ratio. To extend the range a second slug is introduced, at an adjustable distance from the first. When the two slugs are separated an integral number of half wavelengths, the standing wave introduced by one is exactly compensated by the other, in the manner of a one-to-one transformer. When the separation is an odd number of quarter waves, however, the combination introduces a standing-wave ratio equal to K^2. This results from the fact that the mismatch of one slug is then multiplied by the mismatch of the other. Thus if the dielectric constant is 3.16 standing-wave voltage

ratios up to $(3.16)^2 = 10$ can be matched. Polystyrene dielectric can match ratios up to 6.8 ($K = 2.6$).

The magnitude of the ratio is controlled by adjusting the spacing between the slugs (between quarter-wave and half-wave separations). The phase of the standing wave, relative to the standing wave existing in the line external to the tuner, is adjusted by moving both slugs as a unit. The same physical action, while not so simple to analyze, occurs when dielectric slugs are introduced in waveguides. The principal disadvantage of the dielectric slugs is their tendency to break down under voltage stresses in high-power transmission systems.

Metal slugs may also be used. In the coaxial case the impedance of the line containing the slug has the value

$$Z_s = Z_0 - \frac{138 \log_{10} (r_2/r_1)}{r_1} \text{ ohms} \tag{510}$$

where r_2 and r_1 are the outer and inner radii of the cylindrical slug. A quarter-wavelength slug of this impedance introduces a standing-wave voltage ratio of

$$k = \sqrt{\frac{1}{1 - Z_s/Z_0}} \tag{511}$$

Two slugs, spaced an odd number of quarter wavelengths, can match a ratio equal to the square of this quantity or

$$k = \frac{1}{1 - Z_s/Z_0} \tag{512}$$

This expression shows that large values of k may be matched by the use of large values of Z_s. The larger values of Z_s, however, are associated with thick slugs and consequent tendency toward dielectric breakdown between slug surface and coaxial conductors. The adjustment is critical when large values of Z_s are used. Metal slugs may also be used in waveguides.

A combination of two coaxial stub tuners, known as a "power divider," may be used to divide the power in any proportion between two lines. The construction is shown in Fig. 466. The two output terminations are matched to their respective outgoing lines or guides. The position of one stub plunger differs from that of the other by a quarter wavelength, and the spacing between the stubs is a half wavelength. Under these

conditions the output terminations remain matched regardless of the setting of the plungers. Suppose first that the plunger of the shorter stub is set at a quarter wavelength. Then this stub constitutes a conventional quarter-wave stub support, and energy passes to the associated output termination (number 1) without loss. The other stub plunger, being set a quarter wave further out, is a half-wave stub and constitutes a short circuit. Hence, no energy passes to its output termination (number 2). All the energy then passes to output termination 1.

If the plungers are extended an additional quarter wavelength, the two stubs exchange roles, the open circuit becomes a short circuit, and vice versa. The energy then passes completely

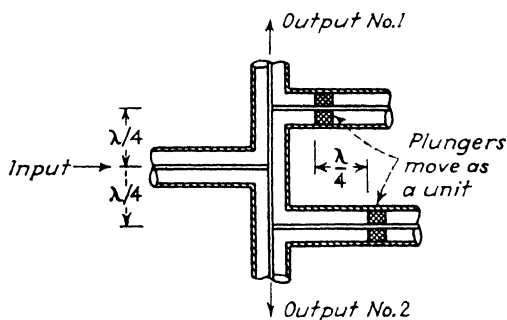



FIG. 466.—Coaxial power divider.

to output 2. At intermediate positions of the plungers, the energy is divided in intermediate proportions, varying as the square of the cosine or sine of the plunger position in electric degrees. Equal division of the power occurs when the plunger settings are, respectively, three-eighths and five-eighths wavelengths from the common line, and at successive odd numbers of eighth wavelengths. The maximum attenuation in the cutoff branch is about 50 db in practical power dividers.

A somewhat similar device, intended to separate the power associated with the direct wave from that associated with the reflected wave, is the directional coupler. A typical form, applied to rectangular waveguide, is shown in Fig. 467. Two openings in the flat surface of the guide, separated a quarter wavelength in guide, lead to a chamber one end of which contains an absorbing load and the other end a power measuring device. The direct wave, passing from left to right, enters the

two apertures and recombines at aperture 2 in constructive interference, since the two portions of energy travel the same distance. Hence this energy passes on to the power measuring device. None of the direct energy passes to the absorbing load, however, since at aperture 1, any d-w (direct-wave) energy traveling to the left is canceled by energy traveling in the same direction from aperture 2, which, having traveled a half wavelength further, interferes destructively at aperture 1. Hence all the d-w energy passing through the two apertures goes to the right toward the power-measuring device, none to the left.

Consider now the reflected wave in the main line, traveling from right to left. Energy from this wave, traveling to the left

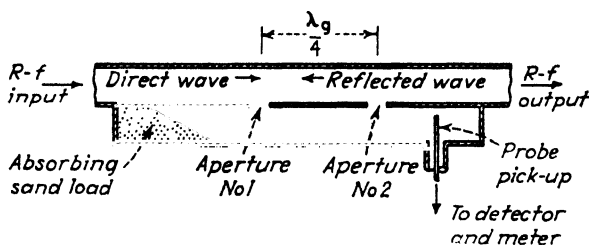



FIG. 467.—Directional coupler.

and entering the apertures, interferes constructively at aperture 1 and passes on to the absorber where it is totally absorbed. Reflected-wave energy, passing through the apertures and traveling to the right, suffers destructive interference at aperture 2 and hence does not pass on the power measuring device. By this means the power measured is restricted to the direct wave only. The directional coupler is used to sample power in a transmission line just before the radiator since the power-measuring device measures only the power actually radiated from the antenna. Reflected signals of any sort, including external reflections as well as those due to mismatch at the antenna itself, are not measured.

220. Measurement of Q.—The measurement of the Q of a resonator is carried out by plotting a resonance characteristic (power vs. frequency) and noting the percentage width $2\Delta f/f_0$ of the resonance curve at the half-power level (3 db down from the response at resonance). The derivation of this relationship ($Q = f_0/2\Delta f$) is given in Sec. 82, Eqs. (264) to (271) inclusive. In practice the measurement may be performed on a point-by-point basis, adjusting the frequency applied to the device under

test and noting the response on a calibrated crystal detector. Since the current output of a crystal closely represents the power level, the half-power points are identified as half-current points on the crystal meter. The essential elements of point-by-point Q measurement are shown in Fig. 468. The bench oscillator is adjusted to successive frequencies, each of which is measured by

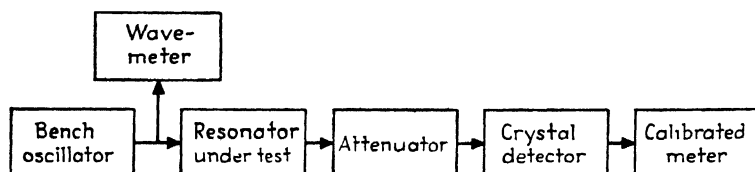



FIG. 468.—Block diagram of point-by-point Q-measurement system.

the wavemeter. The response to the r-f device is measured by the crystal detector and meter. Care is taken to avoid reflected waves in the measurement circuit, by the insertion of attenuator pads, or lengths of "lossy" cable, between the device under test and the crystal.

In the system shown, the device is measured by the transmission method, which is applicable only to resonators which

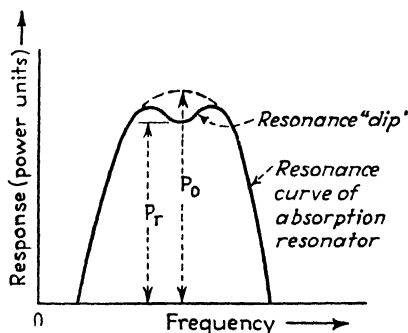



FIG. 469.—Measurement of Q from absorption measurements.

have two terminations. Single-termination devices must be measured by the absorption method. The resonance curve in this case has a dip at the center, corresponding to the point of maximum absorption by the device under test, as shown in Fig. 469. The power level P_r at the center of the dip, compared with the power level P_0 obtained with the device removed from the circuit, indicates the resonator conductance G in accordance

with

$$\frac{P_r}{P_0} = \frac{1}{1 + 1/G + 1/(4G^2)} \tag{513}$$

From the value of G so computed, the half-power amplitude P_{hp} may be derived from

$$\frac{P_{hp}}{P_0} = \frac{1}{1 + 1/(2G) + 1/(8G^2)} \tag{514}$$

The frequency at which the level P_{hp} is found on the resonance curve (Fig. 468) is separated from the resonance frequency f_0 by one-half the quantity $2\Delta f$, from which $Q = f_0/2\Delta f$ is obtained.

The Q of single-termination devices may also be measured by noting the standing-wave ratio in a slotted section adjacent to the device at resonance and at discrete frequencies either side of resonance. The standing-wave ratio at the half-power level is

$$k_{hp} = \frac{k_0 + 1 + \sqrt{k_0^2 + 1}}{k_0 + 1 - \sqrt{k_0^2 + 1}} \tag{515}$$

The frequency separation on the curve between the two levels k_{hp} is $2\Delta f$ in $Q = f_0/2\Delta f$.

The point-by-point method is tedious and is moreover limited in accuracy when high values of Q are to be measured because it is difficult to read the wavemeter accurately enough to plot the narrow resonance curve with precision. When high values of Q are to be measured, a frequency calibration may be introduced from an auxiliary oscillator having a sideband output frequency at a known difference from the carrier value. This is done by modulating an auxiliary tunable bench oscillator with a l-f (say 1 to 40 megacycles) oscillator of the type used as l-f signal generators. The auxiliary oscillator output then contains a carrier frequency and two sidebands separated from the carrier by the frequency of the l-f modulation. The modulated auxiliary oscillator is tuned to such a frequency that one of the sidebands falls directly on the carrier of the main bench oscillator. Frequencies above and below the main resonance are then measured by noting the change in the setting of the l-f modulating oscillator. A cathode-ray indicator is required to indicate the coincidence between the auxiliary sideband and the main carrier signals.

Since a cathode-ray tube is required, use is made of the cathode-ray tube to indicate the resonance curve as well, as described in the next section.

221. Spectrum Analyzer.—A measurement device of wide utility in r-f measurements is the spectrum analyzer, a block diagram of which is shown in Fig. 470. Essentially the spectrum analyzer is a radar receiver, in which the local oscillator is frequency modulated by a sawtooth wave applied to the repeller electrode. The same sawtooth drives the horizontal deflection

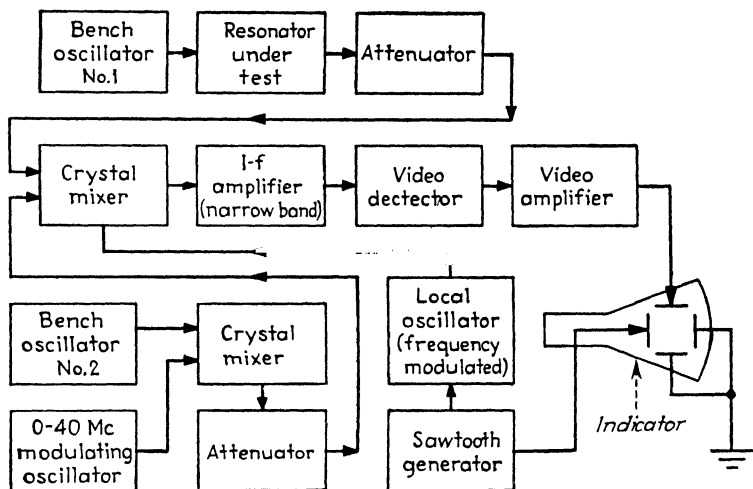



FIG. 470.—Block diagram of spectrum analyzer.

of an oscilloscope. Input signals, applied through an attenuator, are mixed with the local oscillator in a crystal. The intermediate frequencies so developed are passed through a series of i-f stages. These stages differ from those of a conventional radar receiver in that they pass a narrow band (about 150 kc at the half-power level). The i-f stages are tuned to 30 megacycles. A second detector and succeeding video amplifiers apply the signal to the vertical deflection plates of the c-r tube. Also provided are an auxiliary c-w oscillator and l-f modulating oscillator for measuring frequency differences by the method outlined in the previous section.

The spectrum analyzer may be used to indicate the spectrum produced by any r-f device. It was originally devised to examine

the spectral content of pulses produced by cavity magnetrons. The magnetron output, suitably attenuated, is applied to the input. The f-m local oscillator (swinging over a range of 100 to 200 megacycles) is adjusted so that its center frequency is separated 30 megacycles from the magnetron carrier. Then, as the local oscillator swings in frequency, it generates successively a 30-megacycle intermediate frequency for the carrier and for each sideband in the magnetron output above and below the carrier. These i-f components are demodulated and applied to the c-r tube, where they appear plotted against frequency in their proper relative amplitudes (the amplitude is proportional to power, since the crystal detector squares the input voltage).

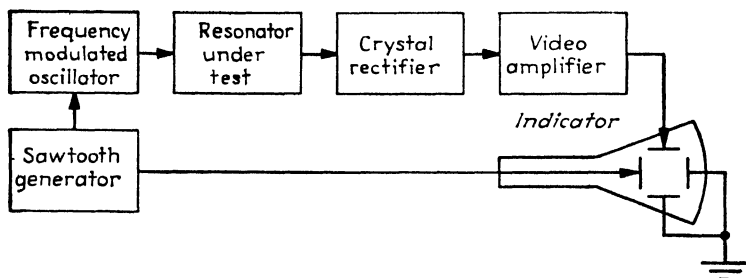



FIG. 471.—Simple visual resonance indicator.

This original function of spectrum analysis has been broadened to include other r-f measurements. When a single-frequency (unmodulated) c-w signal is applied to the input (from the auxiliary bench oscillator, for example), the i-f amplifiers are excited as the f-m local oscillator swings through a value separated 30 megacycles from the input signal. If the local oscillator has constant output, the output appearing on the c-r tube is a pulse whose waveform is the response curve of the i-f amplifier. The response of typical wideband radar i-f amplifiers may thus be studied by insertion in place of the narrow-band i-f amplifier.

Pulses so formed may be calibrated by marker signals. As shown in Fig. 470 an auxiliary c-w oscillator, tunable in the range over which the f-m local oscillator swings, is modulated by a variable l-f signal (1 to 40 megacycles). The auxiliary carrier and sidebands are applied through an attenuator to the spectrum analyzer input. The auxiliary carrier is then tuned to a point just outside the swing region of the f-m local oscillator, but

close enough so that the sideband signal beats with the local oscillator signal and hence appears as a pulse on the c-r tube. The position of this pulse can be adjusted by changing either the local oscillator center frequency or the auxiliary carrier frequency. The sideband signal pulse may be brought into exact synchronism with any other c-w signal present by superimposing the two pulses and noting the position at which the superimposed signals develop a fuzzy appearance due to the beat note between them. Frequency deviations each side of this coincidence point can be measured accurately by shifting the l-f modulating signal, care being taken to keep local oscillator and auxiliary carrier frequency stable during the measurement.

Measurement of the Q of transmission-type resonant cavities can be carried out by passing the local oscillator f-m signal through the cavity and mixing the output with a fixed c-w signal, spaced 30 megacycles to produce the intermediate frequency. A wide-band i-f amplifier should be used in this case (i-f gain constant over the region covered by the frequency swing). A simpler method of producing the same result is shown in Fig. 471. Here the output of the f-m oscillator is passed directly through the cavity under test and thence to a crystal detector and amplifier, which applies the signal to the vertical plates of the oscilloscope.

For measurement of very high Q resonators, the point-by-point system may be used in conjunction with the spectrum analyzer. A tunable c-w oscillator applies a signal to the resonator and thence to the input of the analyzer. By the mechanism previously described, a pulse appears on the c-r screen, the amplitude of which corresponds to the amplitude of oscillation in the cavity at the frequency then in use. By adjustment of the auxiliary oscillator, the amplitude may be measured, one point at a time, for various applied frequencies, each of which is measured accurately by the deviation method, using the marker circuit. In this manner, accurate determinations of Q as high as 100,000 may be made.

Bibliography

GAFFNEY, F. J.: Microwave Measurements and Test Equipments, *Proc. I.R.E.*, **34**, 775 (October, 1946).

GREEN, FISHER, and FERGUSON: Techniques and Facilities for Microwave Radar Testing, *Bell System Tech. J.*, **25**, 435 (July, 1946).

WILSON, SCHRAMM, and KINZER: High Q Resonant Cavities for Microwave Testing, *Bell System Tech. J.*, **25**, 408 (July, 1946).
EARLY, H. C.: A Wide-band Wattmeter for Waveguide, *Proc. I.R.E.*, **34**, 803 (October, 1946).
———A Wide-band Directional Coupler for Waveguide, *Proc. I.R.E.*, **34**, 883 (November, 1946).
Radiation Laboratory Report, "Microwave Technique, as of May 1943." No. T-13, M.I.T. Radiation Laboratory, 1943.
GROSS, E. E., Jr.: Coaxial Butterfly Circuits, *Electronics*, **19** (4), 156 (April, 1946).

Index

A

Absorption, of signals by atmosphere, 279
Absorption cross section, 53
  of antenna, 254
Absorption measurement of Q, 621
Absorptive properties of directive radiators, 254
Accumulator, diode, 334
Admittance, characteristic, 145
Admittance circle diagram, 149, 158
A-f-c circuit, sawtooth, 522, 532
Aided tracking system, 584
Alford loop, 411
Amplidyne drive, 570
  control amplifier for, 571
Amplification, video, 338, 516
Amplification method, choice of, 503
Amplifier, deflection, 576
  limiting, 320
  paraphase, 551
  pulse-excited, dynamic action of, 85
  radio-frequency, 412
  video (*see* Amplification, video)
Amplifier design, intermediate-frequency, 507
Amplitude spectra of various pulses, 100
Amplitude spectrum of carrier, 126
  of repetitive rectangular pulse, 99
  of transient rectangular pulse, 103
Amplitude system function, 110
Analyzer, spectrum, 623
AN/MPG-1 radar, 491–495, 529–533, 584–589
  a-f-c circuit, 532
  azimuth markers in, 587
  indicator used in, 584
AN/MPG-1 radar, marker circuits of, 586
  ppi system in, 585
  type B indicator of, 586
AN/MPG-1 driver, 492
AN/MPG-1 i-f amplifier, stagger-tuned, 531
AN/MPG-1 receiver, 530
AN/MPG-1 sensitivity-time-control circuit, 532
AN/MPG-1 transmission system, 494
Anomalous propagation of radar signals, 270, 298
Antenna, absorption of, 254
  horn, 198
  J-type, 410
    (*See also* Radiators)
Antenna coupling circuit, 413
Anti-t-r, switch, 398
AN/TPS-3 radar, 433–439, 478–484, 525, 574–577
  auxiliary timer circuits of, 436
  dipole feeds of, 481
  indicator of, 574
  modulator of, 478
  radiator of, 480
  receiver of, 525
  r-f oscillator of, 478–479
  scanner of, 576
  signal circuits of, 577
  t-r system of, 483
Apertures, capacitive, 202
  inductive, 202
Appleton, Sir E., 6, 61
Area, echo, of target, 283
Array, angular distribution of electric field, 239
Arrays, dipole, directive properties of, 233–235

Arrays, dipole, gain of, 238
physical forms of, 244
discrete half-wave elements, 245
mattress, 246
gain of, 246
single-sided radiation from, 240
Astigmatism control, 544
Asymmetrical dipole, 406
Atmosphere, absorption of signals by, 279
Atmospherics, 56
Attenuation, in circular waveguide, 194
in coaxial lines, 171
in waveguides, 191
rectangular, 193
Attenuation constant, 140
for rainfall, 280
Attenuation losses in dielectric, 173
Attenuators, 600–604
flap-type, 603
plug-type, 604
strip-type, 603–604
waveguide-type, 601
Automatic brightness control, 545
Automatic frequency-control circuit, 520
Automatic tracking system, SCR-584, 529, 582
Auxiliary circuits for cathode-ray indicators, 543
Available power, 131
Azimuth, 25
Azimuth markers, system for generating, 587

B

Back scattering, 281
Background returns, 46
Balanced to coaxial transformations, 378
Bandpass characteristics, single- and double-tuned, 514
Bandpass curves of i-f amplifier, 511
Bandwidth, 56, 301, 507
criterion for high resolution, 129
critical value of, 129
Bandwidth, for maximum signal-to-noise ratio, 127
of receiver, 24, 508
of system, effective, 55
Barnes, C. B., 310
Barnett, M. A. F., 6
Barrow, W. L., 181, 217, 218, 290
Beacon, radar, 409
Beam, conical, 239
radiated, specifications of, 294
Beam width, 52–54, 295
of mattress array, 247
Beams, formation of by electromagnetic lenses, 256
Bench oscillators, 591
Bends in waveguide, 204
Beringer, R., 279
Berkner, L. V., 62, 310
Biconical horn, 411
Blair, Major W. R., 7
Blocking oscillator, 317, 336
synchronized, 324
waveform of, 318
Blocking-oscillator driver, line-controlled, 473
Blooming, 544
Bolometer probe, 596
Bootstrap dircuit, 346
Boostrap driver circuit, 471
Box, transmit-receive (t-r), 395
Brainerd, J. G., 217, 290
Breit, G., 7
Bridge, thermistor, 594
Brightness-control, 544–545
automatic, 545
Bunches in klystron, 419
Bunching parameter, 422
Byrnes, I. F., 310

C

Calibration system, microwave-frequency, 606
Calorimeter, water, 598
Capacitance, transient response of, 68
Capacitive apertures, 202
Carrier, amplitude spectrum of, 126

Carrier, pulse-modulated, transmission of, 124
Carrier wave, envelope of, after passage through filter, 126
Carson, J. R., 218
Cascade phosphor, 539
Cascaded stages, i-f, 512
Cathode in magnetrons, 390
Cathode-follower circuit, 348–349
Cathode-ray indicator, 536
  auxiliary circuits for, 543
Cathode-ray-indicator tube, 537
Cavities, for disk-seal tube, 414
  resonant, 205
  transmit-receive, 401
  wavelength in, 213
Cavity magnetron, 378
  internal action of, 381
  typical characteristics, 392
Cavity resonators, excitation of, 217
  modes in, 211
Cavity wavemeters, 608
Characteristic admittance, 145
Characteristic impedance, 142
  construction for finding, 156
Choke joint, 364
Chu, L. J., 181, 268, 290
Circle diagram, admittance, 158
  applied to waveguide, 200
  impedance, 149
  of loss-free line, 154
  Smith-type, 160
Circuit, differentiator, 322
Circuit elements, transient response of, 64
    pairs of, 70
Circuits, equivalent, of cathode-follower, 349
    of ring oscillator, 376
  for generating trapezoidal wave, 553
  oscillating, 205
  pulse, basic, 311–358
  range-marker, 438
  selection of, 302
Circular guides, transverse electric modes, 189
  transverse magnetic modes, 190
Circuit Q, 206
Circular scanning, 26
Circular waveguide, modes in, 188
Clamping circuit, 356, 574
Coaxial to balanced transformations, 378
Coaxial butterfly circuit, 593
Coaxial inner conductors, support of, 359
Coaxial joints, choke flange, 363
Coaxial line, 137
  attenuation in, 171
  fittings, 359
  standing-wave detector for, 613
Coaxial resonator, geometry of, 215
  reentrant, 216
Coaxial rotating joint, simple capacitive, 363
Coaxial stub supports, 360
Coaxial wavemeters, 607
Coaxial-waveguide connectors, 366
Cole, A. D., 310
Colpitts oscillator, 371
Colton, R. B., 61, 310
Compensation, high-frequency, 343
  low-frequency, 345
  of video amplifiers, 343
Components, radio-frequency, 359–430
  of SCR-268 radar, 477
  superheterodyne, 415
Condon, E. U., 218
Conical beam, 239
Conical scanning, 29
  radiators for, 405
Connectors, coaxial-waveguide, 366
Continuous-wave radar system, 15
Control amplifier, for amplidyne drive, 571
  for Ward-Leonard drive, 569
Control transformer, selsyn, 566
Converter, crystal frequency, 425
Cordinates, specifying target position, 25
Corner reflector, 284
Corners, waveguide, 204
Cosecant-squared distribution, 253
Coupler, 614

Coupler, directional, 620
Coupling circuit, antenna, 413
Coupling methods, intermediate-frequency amplitude, 510
  to transmission lines, 377
Coverage diagram, envelopes of lobes, 269
  lobes in, 269
  vertical, 263, 297
Crystal, noise contribution of, 428
Crystal calibration, silicon, 598
Crystal frequency converter, 426
Crystal rectifier, equivalent circuit of, 427
Crystals, silicon, 426
Cubical resonator, geometry of, 212
Cutoff, gradual, 118
  grid-current, 86
  grid-voltage, 80
  sharp, 118
C-w radar to determine height, 16
Cylinder, paraboloidal, 252
Cylindrical echo boxes, 609
Cylindrical resonators, 213
  geometry of, 214

D

Data-transmission systems, 563
Davis, R. J., 589
Deflection, definition of, 32
  direction-sweep, circuit for generating, 561
Deflection amplifiers, 576
Deflection baseline, disortion of, 560
Deflection patterns for various types of indication, 545
Deflection systems, 540
  electric, 540
  magnetic, 541
Deionization time, 397
Delay multivibrators, 355
Detection, range of, effect of precipitation on, 282
  video, 516
Detection distance, maximum, 19
  minimum, 18
Detector, for coaxial lines, 613
  diode, 92
  frequency-modulation, 520
  video, 518
Detectors, standing-wave, 610
  for rectangular guide, 612
DeWitt, J. H., Jr., 62
Diagram, Rieke, 393
  circle, 158
  vertical coverage, 263
Dielectric, losses in, 173
Differentiator circuit, 322
Dimensions, optimum, for power transfer, 173
  of waveguides, 195
Diode, accumulator, 334
  transient operation of, 91
Diode circuit, negative limiting, 91
  positive limiting in, 91
Diode clamping circuit, 356
Diode detector, 92
  polarities, 517
Diode d-c restorer, 357
Dipole, 53
  asymmetrical, 406
  computing polar-diagram array, 234
  electric, 219
  electric-field strength, polar diagram of, 224
  fields surrounding, 223
  gain of, 226
  infinitesimal oscillating, properties of, 220
  power radiated from, 224
    polar diagram of, 225
Dipole array, gain of, 238
  directive properties of, 233
  directivity of, 235
  physical forms of, 244
Dipole director, 241
Dipole elements, excitation of, 245
Dipole feeds of AN/TPS-3 radar, 481
Dipole radiation resistance, 226
Dipole target, 284
Direct wave, 139
D-c component, 560

D-c reinsertion, 545
D-c restorer, diode, 357
Directional coupler, 620
Direction-sweep, circuit for generating, 561
Direction-sweep potentiometer, taper for, 563
Directive properties of dipole arrays, 233–235
Directive radiators, absorptive properties of, 254
Director, dipole, 241
Discharge tube, constant-current, 551
Discriminator circuit, 520
Disk-seal triode, 369
Disk-seal tube, cavities for, 414
Distortion, effects of amplitude and phase, 123
  introduced by type B indicators, 34
  introduced by type C indicator, 35
  waveform, 94
Distortionless type B presentation, 589
Distribution, cosecant-squared, 253
Divergence factor, 268
Divider, power, 619
Division, frequency, 335
Dolph, C. L., 242
Doppler effect, 15, 287
  geometry of, 288
  magnitude of frequency shift due to, 289
Double-resonator klystron, 417, 593
Double-stub tuners, 615–616
  circle diagram of, 615
Double-tuned bandpass characteristics, 514
Drive, amplidyne, 570
    control amplifier for, 571
  speed-control, Ward-Leonard, 568
Driver, AN/MPG-1, 492
  line-controlled blocking-oscillator, 473
  saturable-core reactor, 474
  SCR-584, 485
Driver circuit, bootstrap, 471
Driver circuits, 470
DuBridge, L. A., 61
Duct, 272
Duplex switching systems, 395
Duty cycle, 20
Dynamic action, of positive limiting, 87
  of pulse-excited amplifier, 85
Dynamic characteristics, 82

E

Echo, target, 51, 59
Echo area of target, 50, 283
Echo boxes, 609
Echo-box trace, 608
Echoes, terrain, 47
Edson, W. A., 217, 290
Eighth-root law, 265
Electric deflection, 540
  radial, 542
Electric dipole, 219
Electromagnetic lens, 256
Electron gun, 537
Elevation, 25
Energy storage, 209
Envelopes of coverage diagram, 269
Equation, Hartree, 386
  radar, free-space, 48
Error signal, 583
Even function, 97
Excitation, of cavity resonators, 217
  of waveguides, 197

F

Feed, linear dipole, 252
  nutating, 407
  for paraboloidal reflectors, 249
  for paraboloids, waveguide, 251
Feedback voltage, 372
Ferguson, J. G., 625
Fields surrounding dipole, 223
Figure, noise, 133
Filter, gradual cutoff, 119
  ideal low-pass, phase system function of, 116
    pulse transmission by, 120
    rectangular pulse by, reproduction of, 122

Filter, ideal low-pass, system functions of, 115
sharp cutoff, 119
Filter sections, design formulas for, 79
Fink, D. G., 17, 61, 310, 393
Fisher, H. J., 625
Fisk, J. B., 381
Fixed loads, 600
Flap-type attenuator, 603
Flat metallic target, 284
Flat reflecting earth, propagation over, 260
Focusing, phase, 387
Foster, J. S., 407
Fourier integral, 101
Fourier series, 94–95
Frank, N. H., 268
Free-space radar equation, 48
Frequency, calibration system, microwave, 606
control circuits, 520
employed in radar, 10
Frequency converters, heterodyne, 425
Frequency dividers, 336
multivibrator as, 335
Frequency multiplier, multivibrator as, 337
Frequency-modulation detector (discriminator), 520
Friis, H. T., 499
Functions, of indicators, 534
of scanners, 534

G

Gaffney F. J., 625
Gain, definition of, 52
of dipole, 226
of dipole array, 238
of mattress array, 246
of paraboloid, 249
of waveguide radiator, 231
Galambos, R., 4
Gating of i-f amplifier, 515
Gebhard, L. A., 7
Generators, radio-frequency, 366
saturable core, 329
Geometry, of Doppler effect, 288
of propagation over flat earth, 260
of propagation over spherical earth, 267
of rectangular waveguide, 185
Ginzton, E. L., 423
Green, E. I., 625
Greene, F. M., 290
Grid-current cutoff, 86
Grid-voltage cutoff, 80
negative limiting by, 83
Griffin, D. R., 4
Ground return, 46
Grounded-grid circuit, 414
Group velocity, 179–180
Guillemin, E. A., 217
Guillemin line, 326

H

Half-wave elements, arrays of, 245
Half-wave radiator, polar diagram of, 228
power radiated by, 228
properties of, 227
Hansen, W. W., 217
Harrison, A. E., 423
Hartley oscillator, 371
Hartree equation, 386
Heaviside, O., 6
Helical scanning, 27
Helmholtz coil, 353
Hertz, Heinrich, 5
Heterodyne converters, 425
High-speed scanners, 407
rotating-feed, 408
Historical survey of radar, 5
Holdam, J. V., 310
Horizon, optical, 266
radio, 272
Horn, biconical, 411
Horn antenna, 198, 231
Horn radiator, pyramidal, 233
Hunting, 568

I

I-f, cascaded stages, 512
I-f amplifier, 531

I-f amplifier, bandpass curves of, 511
gating of, 515
interstage coupling in, 509
stagger tuning of, 513
I-f pickup for tuning indication, 519
I-f-f transponder, 409
Impedance, characteristic, 142
construction of Smith, 162
normalized, 150
presented by radiator in scanning, 403
specific wave, 198
spiral, 153
terminating, 142
Impedance circle diagrams, 149
Impedance matching in waveguides, 200
Index of refraction, 272
modified relative, 273
Index curves, modified, measurement of, 275
typical, 276
Indication, deflection patterns for, 545
ppi, 44
(*See also* Indicators)
Indicators, 534
AN/MPG-1 radar, 584
AN/TPS-3 radar, 574
cathode-ray, 536
auxiliary circuits for, 543
functions of, 534–535
plan-position, 36
deflection system for, 548
SCR-268, 572
SCR-584, 577
plan-position, 578
specifications for, 32, 294, 301
sweep systems of, plan-position, 554
type A, 32
deflection system for, 546
modifications of, 37
type B, 33
deflection system for, 546
distortion introduced by, 34
type C, 35
deflection system for, 547
Indicator, type C, distortion introduced by, 35
type D, 40
type E, 39
type F, 39
type G, 39
type H, 39
type I, 40
type J, 37
deflection system for, 549
SCR-584 radar, 448
type K, 38
type L, 38
type M, 38
type N, 39
type R, 39
types of, summary, 40
Indicator tube, cathode-ray, 537
Inductance, transient response of, 67
Inductive apertures, 202
Inductive output tube, 415
Inglis, A. F., 316
Input impedance, of loss-free lines, 147
transmission line, 146
Integrator circuit, 334
Interference, man-made, 56
Intermediate frequency, choice of, 505
Intermediate-frequency amplifier, coupling methods, 510
design of, 507
Interstage coupling in i-f amplifiers, 509
Isotropic radiator, 219

J

Jamieson, H. W., 368
Johnson power meter, 599
Joint, bullet and flange, 362
rotating, in coaxial lines, 362
waveguide, 364
J-type antenna, 410

K

K band, 135
Keep-alive electrode, 400

Kennelly, A. E., 6
Kerr, D. E., 258
Kiebert, M. V., 316
King, A. P., 279
Kinzer, J. P., 626
Klystron, 415, 417
  bunches in, 419
  double-resonator, 417, 593
  multiplier, 595
  reflex, 417, 592

L

L band, 135
LC response, 74
Lenses, electromagnetic, 256
Leverenz, H. W., 539
Lewis, F. D., 290
Lewis, W. D., 407
Lighthouse tube, 369
Limited wave, time of rise of, 320
Limiters, deformation, 319
Limiting amplifier, 320
Line, coaxial, 137, 171
  Guillemin, 326
  half-wavelength, properties of, 148
  loss-free, 145
    input impedance of, 147
  open wire, 137, 170
  pulse-forming, 325
  transmission, 137
Linear dipole feed, 252
Linear sweeps, 549
Line-controlled blocking-oscillator driver, 473
Load, plug, 605
    sand, 605
Load line method, 82
Lobe, main, 236
  side, 237
    reduction of, 242
  vertical elevation of, 405
Lobe shifting by phase relationships, 243
Lobe switching, 31, 403
Local oscillator, 415
Local-oscillator circuits, 423
Long radiators, linear, 229
  polar diagram, 229
Loop, Alford, 411
Loss, reflection, 164
Loss-free lines, 145
  circle diagram of, 154

M

M curves, relation of, to weather phenomena, 278
McArthur, E. D., 368
Magnetic deflection system, 541
Magnetron, cavity, 378, 417
    internal action of, 381
    typical characteristics, 392
  cathode in, 390
  electron paths in, 383
  equivalent circuit of, 395
  internal structure of, 379
  magnet structures, 391
  operating characteristics of, 388, 391
Main lobe, 236
Marchetti, J. W., 310
Marker circuits of AN/MPG-1, 586
Markers, azimuth, 587
Mattress arrays, 246
  beam width of, 247
  gain of, 246
Maxwell, Clerk, 5
Maxwell equations, 183
McGrath, S., 310
Mead, S. P., 218
Measurements, radio-frequency, 590
Microwave mixers, 430
Micher, W. W., 218
Military characteristics, 293
Military radars, specifications, 302
Mixers, 425
  microwave, 430
  pencil, 429
  pot, 429
Modes, in cavity resonators, 211
  in waveguides, choice of, 195
  in circular waveguide, 188
  methods of exciting, 198
  principal, 199

Modified index curves, measurement of, 275
typical, 276
Modified relative index of refraction, 273
Modulated sawtooth waves, circuit for generating, 558
Modulation, of sawtooth waves for ppi, 557
of test oscillators, 591
Modulator, 486, 493
of AN/MPG-1 radar, 493
of AN/TPS-3 radar, 478
functions of, 458
high vacuum tube, 463
inductance as storage element, 470
peaking inductance in, 466
pulse-forming line, 467
rotary-spark-gap, 462
of SCR-268 radar, 475
of SCR-584 radar, 486
switching element, 461
thyratrons used in, 464
Modulator circuits, basic, 459, 465
d-c charging, 468
elements, types of, 460
hard-tube, 465
synchronous, 469
Modulator high-voltage power supplies, 460
Mofenson, J., 62
Moreno, T., 193
Mueller, G. E., 279
Multiplier klystron, 595
Multivibrator, 313
delay, 355
direct-coupled, 330
as frequency divider, 335
as frequency multiplier, 337
symmetrical, waveforms, 315
synchronized, 324
time constants associated with, 316
unsymmetrical, 316

N

Negative limiting, in diode circuit, 91
by grid-voltage cutoff, 83
Neper, 172
Networks, single-terminal-pair, 104
two-terminal-pair, 113
system functions of, 112
Noise, 54, 128
concept of, in radar, 130
in pulse systems, 127
in receivers, 499
Noise contribution of crystal, 428
Noise figure, 55, 133, 301
due to mismatching, improvement of, 133
over-all, 500
receiver, 24
Noise power, available, 55
Noise sources, physical nature of, 502
Nondirectional antenna, triple-dipole, 411
Nondirectional radiators, 409–410
Nutating feeds, 407

O

Off-axis radiation, 406
Offset radiator for conical scanning, 406
One-shot multivibrator, 331
Open-wire line, 137
Optical horizon, 266
Oscillating circuits, 205
Oscillator, bench, 591
blocking, 317
Hartley, 371
internal-circuit, 375
local, 415
circuits of, 423
relaxation, 312
ring, 375
equivalent circuits of, 376
shock-excited, 338
sine-wave, 312
transmitting, 366
triode, external-line, 367
tuned-grid tuned-cathode, 374
tuned-plate tuned-grid, 373
waveform of, 318
Oscillator circuits, Colpitts, 371
ultrahigh-frequency, 370

P

P band, 135
Paraboloid, distorted, 253
  gain of, 249
  waveguide feeds for, 251
Paraboloid polar diagrams, 250
Paraboloidal cylinder, 252
Paraboloidal reflectors, 247
  feeds, 249
  geometry of, 248
Paraboloidal sections, 251
Parallel-line tuned circuit, 369
Parallel-tuned circuit, 205
Parameters of transmission lines, 170
Paraphase amplifier, 551
Path difference, 261
Pattern-propagation factor, 258
Peak power, 300
Peaking inductance in modulator, 466
Pencil mixer, 429
Performance requirements, 293
Permanent echoes, 43
Peterson, E. F., 368
Phase, lobe shifting, 243
Phase constant, 140
Phase focusing, 387
Phase shift, of sinusoidal sources, 352
Phase shift, generator-type, 353
Phase system function, 110
Phase velocity, 141, 179, 181
Phosphors, 539
  cascade, 539
π-mode, r-f electric field, 385
Plan-position indicator, 36
  deflection system for, 548
  SCR-584, 578
  sweep systems, 554
Plug attenuator, 604
Plug load, 605
Polar diagram, half-wave radiator, 228
  dipole, 224
  of long radiators, 229
  paraboloid, 250
  of power radiated from dipole, 225
Polarities, diode detector, 517
Positive limiting, in diode circuit, 91
  dynamic action of, 87
Pot mixer, 429
Potential, scalar, 221
  vector, 221
Power, available, 131
  average, 21
  density of, 49
  measurements of, in transmission lines, 164
  output of, 416
  peak, 21
  radiated, from a dipole, 224
    by half-wave radiator, 228
  transfer of, 166
    from generator to load, 167
    optimum dimensions for, 173
Power divider, 614
  coaxial, 619
Power loss, associated with standing wave ratio, 169
Power meter, Johnson, 598
Power-measuring devices, 593
Poynting vector, 191, 255
Ppi, 36
  (*See also* Plan-position indicator)
  AN/MPG-1, 585
  deflection system, output circuit of, 560
  rotating yoke, 555
Preamplifier, components of, 503
  SCR-584, 527
Precipitation, effect on range of detection, 282
Principal mode, 199
Probe, bolometer, 596
Propagation, 177, 219, 257
  anomalous, of radar signals, 270, 298
  effect of ground reflections on, 23
  off-axis, 259
  over flat earth, 260
  over spherical earth, 265–267
  radio, velocity of, 18
  surface, 59
  transverse magnetic, 182
  velocity of, 141

Propagation, in waveguides, mechanism of, 175
Pulse, 57
physical length of, in space, 19
rectangular, 66, 99, 393
amplitude spectrum of, 99
basic geometry of, 20
nonrepetitive, 102
spectra of repetitive, 97
transient, amplitude spectrum of, 103
skew symmetrical, 98
spectra of various types of, 100
timer, 432
Pulse circuit, auxiliaries, 356
basic, 311
function of, 311
steady-state analysis of, 93
Pulse generation and transmission, principles of, 63
Pulse generators, series resonance in, 328
single-capacitor, 327
synchronous, 328
Pulse rate, 19, 299, 434
variation in, 435
Pulse shape, 300
Pulse specifications, 17, 294, 298
Pulse spectra, determination of, 95
Pulse system, description of, 11
noise in, 127
technical specifications of, 17
Pulse transformers, 350, 467
equivalent circuits of, 350
Pulse transmission, by ideal low-pass filter, 120
by video amplifier, 118
Pulse waveforms, amplification of, 338
Pulse width, 299
Pulse-excited amplifier, dynamic action of, 85
Pulse-forming lines, 325
in cathode circuit, 327
modulator, 467
Pulse-modulated radio-frequency carriers, 124
Pulse-rate division, 335
Pulse-rate multiplication, 335
Pulse-repetition rate, 19, 26
Pulse-shaping, typical sequence of, 335

Q

Q, circuit, 206
measurement of, 620
measurement system, 621
in resonant cavities, 208
Quadrantal capacitor, 353

R

Radar, characteristics of, 293
c-w, 3
definition of, 3
frequencies employed in, 10
historical survey of, 5
military, technical specifications of, 302
noise concept in, 130
origins of, 3
pulse, 3
radio frequencies employed in, 134
technical specifications of representative, 304–309
type AN/APQ-7, 306–307
type AN/APQ-13, 306–307
type AN/APS-3, 306–307
type AN/CPS-1, 304–305
type AN/MPG-1, 304–305
type ASB, 306–307
type Mark 4, 308–309
type Mark 13, 308–309
type SCR-270-B, 304–305
type SCR–584, 304–305
type SCR-602-A, 304–305
type SCR-720, 306–307
type SK, 308–309
type SL, 308–309
type SO-3, 308–309
wartime employment of, 8
Radar beacon, 409
Radar design, 293
Radar detection, methods of, 11

Radar distances, 18
Radar equation, free-space, 48
numerical examples, 58
influence of wavelength on, 51
Radar propagation, geometry of, 49
Radar reflectance, 286
Radar signals, propagation of, 257
anomalous, 270
Radar system, essential elements, 303
Radar target, 40, 282
Radial deflection, 542
Radiation, from arrays, single-sided, 240
off-axis, 406
Radiation Laboratory, 8
Radiation resistance, dipole, 226
temperature of, 132
Radiators, 219, 402, 458, 484
of AN/TPS-3 radar, 480
for conical scanning, 405
definition of, 52
directive, absorptive properties of, 254
half-wave, polar diagram of the electric field, 228
properties of, 227
horn, 231
isotropic, 219
long linear, 229
nondirectional, 409
offset, for conical scanning, 406
polar diagrams, 229
in scanning, impedance presented by, 403
SCR-268, 474
of SCR-584 radar, 488
system, 491
waveguide, 230
Radio frequency, choice of, 297
employed in radar, 134
Radio horizon, 272
Radio-frequency amplifiers, 412
Radio-frequency carriers, pulse-modulated, transmission of, 124
Radio-frequency circuits, 359
Radio-frequency components, 506
Radio-frequency designations, 135
Radio-frequency generators, 366
Radio-frequency measurements, 590
Radio-frequency specifications, 21, 294
Radio-frequency structures, 359
Radio-frequency techniques in radar, 134
Radio-frequency test equipment, 590
Radome, 403
Rainfall, attenuation constants for, 280
Ramo, S., 217, 290
Range, 25
of detection, effect of precipitation on, 282
free-space maximum, 261
Range channel, 526
Range marker, 438
SCR-584, 581
Ratio, signal-to-noise, 413
Ray paths, representation of, 274
RC response, 73
Reactive component, cancellation of, 157
Receivers, 496
AN/MPG-1, 530
AN/TPS-3, 525
bandwidth of, 24
characteristics of, in radar equation, influence of, 54
essential elements of, 498
functions and components of, 496
noise in, 499
noise figure in, 24
SCR-268, 523
SCR-584, 525
simplified schematic of, 528
typical microwave, 525
Receiver specifications, 294, 301
Rectangular echo boxes, 609
Rectangular waveguide, standing-wave detector, 612
transverse electric modes in, 183
various transverse electric modes in, 187
Rectangular pulse, 66, 393
amplitude spectrum of, 99

Rectangular purse, geometry of, 20
nonrepetitive, 102
reproduction of, by ideal low-pass filter, 122
spectra of repetitive, 97
transient, amplitude spectrum of, 103
transmission of, through gradual cutoff filter, 119
through sharp cutoff filter, 119
Reflectance, radar, 286
Reflected signal power, minimum, 54
Reflected wave, 139
Reflecting elements in waveguides, 200
Reflection coefficient, 261
Reflection intervals, 18
Reflection loss, 164
Reflectors, 219
paraboloidal, 247
Reflex klystron, 417, 592
conditions for maximum output, 421
Refraction, index of, 272
Refractive bending of wave paths, 271
Regulated power supply, 357
Reich, H. J., 217
Reinsertion, d-c, 545
Reisman, S. J., 589
Relaxation oscillator, 312
Repeller voltage, 420
effect of, on frequency and power output, 424
Resistance, transient response of, 66
Resolution, bandwidth criterion for, 129
Resonance, methods of indicating, 607
Resonance curves of shunt-tuned circuit, 207
Resonance indicator, simple visual, 624
Resonant cavities, 134, 205
dissipation in, 209
excitation of, 217
modes in, 211
Q in, 208
Resonant wavelength in cavities, 213
Resonator, coaxial, geometry of, 215
reentrant, 216
cubical, 212
cylindrical, 213
spherical, 215
Response, *LC*, 74
RC, 73
RL, 70
steady state of, 65
transient, 65
Return, ground, 46
background, 46
R-f amplifier, typical, 413
R-f electric field, mode, 385
R-f oscillator, 486, 493
R-f stage, design of, 412
R-f switch, rotating joint, 495
R-f transmission system of SCR-584, 487, 490
Ridenour, L. N., 62
Rieke diagram, 393
Ring oscillator, 375
of SCR-268 radar, 476
Ringing time, 609
RL response, 70
Robertson, S. D., 279
Rotary spark gap, 312, 433
modulator for, 462
Rotary transformer, 574, 579
sinusoidal modulation for, 559
Rotating joint, 495
capacitive, 363
in coaxial lines, 362
waveguide, 365
Rotating yoke, ppi, 555
Rotating-dipole, SCR-584 radar, 491
Rotating-feed type of high-speed scanner, 406, 408
Rubenstein, P., 258
Rueger, L. J., 589

S

S band, 135
Salisbury, W. W., 9
Samuel, A. L., 399

Sand load, 605
Sarbacher, R. I., 217, 290
Saturable-core generators, 329
Saturable-core reactor driver, 474
Sawtooth generator, constant-current, 332
Sawtooth wave, 332
  of current, circuit for producing, 552
  in inductive circuit, 333
  for ppi, modulation of, 557
  in resistive-inductive, 333
  synchronized production of, 331
Scalar potential, 221
Scan, definition of, 25
Scanner, 534
  AN/TPS-3, 576
  functions of, 534
  high-speed, 297, 407
  specifications for, 25
Scanning, angular, servomechanism for, 567
  circular, 26
  conical, 29
  definition, 32
  helical, 27
  spiral, 30
Scanning interval, 28
Scanning mechanisms, SCR-584, 580
Scanning rate, 296
  relationships of, 28
Scanning specifications, 294–296
Scattering, back, 281
Schelkunoff, S. A., 218
Schneider, E. G., 62
Schramm, C. W., 626
SCR-268 radar, block diagram of, 456
  components of, 477
  indicator for, 572
  modulator and r-f oscillator, 475
  receiver, 523
  ring oscillator of, 476
SCR-268 radiators, 474
SCR-268 timer, 455–456
SCR-268 transmitter, 474
SCR-584 radar, 484–486
  t-r cavity of, 489
SCR-584 automatic tracking system, 529, 582
SCR-584 driver, 485
SCR-584 radar indicator (type J), 448
SCR-584 plan-position indicator, 578
SCR-584 preamplifier, 527
SCR-584 radiator, 488
SCR-584 range markers, 581
SCR-584 receiver, 525–528
SCR-584 r-f transmission system, 487–490.
  rotating-dipole, for, 491
SCR-584 radar rotating-dipole for, 491
SCR-584 scanning and mechanisms, 580
SCR-584 selsyn data-transmission system, 583
SCR-584 timer, 445–452
  waveforms of the, 447
SCR-720 timer, 441–444
Selsyn, differential, 565
Selsyn control transformer, 565
Selsyn data-transmission systems, SCR-584, 583
Selsyn receiver, 564
Selsyn transmitter, 564
Sensitivity, radio-frequency, 10
Sensitivity-time-control circuit, 532
Servo channel, 526
Servomechanism, 568
  for angular scanning, 567
  relative bearing of, in azimuth, 571
Shock-excited oscillator, 338
Shunt-tuned circuit, resonance curves of, 207
Side lobes, 237
  reduction of, 242
Signal, absorption of, by atmosphere, 279
  minimum discernible, 54
Signal power, minimum reflected, 54
Signal-to-noise ratio, 413
  as function of receiver bandwidth, 508

Signal-to-noise ratio, maximum, bandwidth for, 127
Silicon crystals, 426
Silicon-crystal calibration, 598
Sine-wave oscillator, 312
Single-tuned bandpass characteristics, 514
Sinusoidal modulation for rotary transformer, 559
Skilling, H. H., 175, 217, 290
Skin depth, 172
Slater, J. C., 151, 192, 217, 290
Slotted sections, 611
Slug tuner, dielectric, 617
Smith, P. H., 150
Smith circle diagram, 160
Smith-Rose, R. L., 61
Southworth, G. C., 217, 218
Spark gaps, 433
  modulator in, 461
  rotary type of, 312
Specific wave impedance, 198
Specifications, radar, 293
  indicator, 301
  pulse, 298
  receiver, 301
  scanning, 296
Spectra, pulse, determination of, 95
Spectrum analyzer, 623
Speed-control drive, Ward-Leonard, 568
Sphere, target conducting, 283
Spherical earth, reflecting, propagation over, 265
Spherical resonator, 215
Spiral scanning, 30
Spread voltage, 573
Stagger tuning of i-f amplifiers, 513
  AN/MPG-1, 531
Standing waves, 144, 165
  formation of, 168
Standing-wave detectors, 610
  for coaxial lines, 613
  ratio of, 200
    associated with power loss, 169
  for rectangular guide, 612
Steady-state analysis of pulse circuits, 93
Steady-state computation of output waveforms, 117
Steady-state response, 65
Stevens, W. E., 426
Straus, H. A., 310, 589
Strip attenuator, 603–604
Stub support, broad-band, 361
  coaxial, 360
Superhetrodyne components, 415
Susceptance, cancellation of, 159
  normalized, 203
Sweeps, linear, 549
Switch, anti-t-r, 398
  for microwave systems, transmit-receive, 399
  t-r, microwave, 400
Switching, lobe, 403
Switching systems, duplex, 395
Synchronization equipment, 431
System functions, computation of, 104
  of ideal low-pass filter, 115
  of two-terminal-pair network, 112
  of video coupling connection, 106

T

Taper, for direction-sweep potentiometer, 563
  in resistance winding for deflection, 562
Targets, 219, 282
  conducting sphere, 283
  data requirements of, 294
  definition of, 12
  dipole, 284
  distance of, 51
  distinction between types of, 45
  fixed, 43
  flat metallic, 284
  moving, 43, 287
    detection of, 290
  as scatterer, 41
  specifying coordinates for, 25
  surroundings of, effect of, 286
  types of, 42
Target, contrast, 286
Target echo, 51, 59

Target echo, area of, 50, 283
Taylor, A. Hoyt, 6
Taylor, J. H., 589
Taylor, M., 589
Tees, waveguide, 203
Temperature of radiation resistance, 132
  of space, equivalent, 56
Terman, F. E., 193, 217
Terminating impedance, 142
Termination of waveguides, 197
Terrain echoes, 47
Test equipment, radio-frequency, 590
Test oscillators, modulation of, 591
Thermistor, 594
Thermistor bridge, 594
Thyratron, modulator, 461
Time delay circuits, 352
Timer, 431, 434
  auxiliary, circuits of AN/TPS-3 radar, 436
  master-oscillator, 433
    crystal-controlled, 445
    sine-wave, 439
  SCR-268, 455
  SCR-584, 445–452
  SCR-720, 441
  self-synchronous, 433, 435
Timer pulses, 432
Timer waveforms of SCR-584, 447
  of the SCR-720 radar, 440
Timing equipment, functions of, 431
Timing sources, 312
T-r box, 396
T-r cavities, methods of coupling to, 402
  SCR-584, 489
T-r switch, microwave, 400
T-r system, of AN/TPS-3 radar, 483
  waveguide, 402
T-r transformer cavity, 399
T-r tube, 400
Tracking system, aided, 584
  automatic, SCR-584, 529, 582
Transformation, balanced to coaxial, 378
Transformer, cavity, t-r, 399
Transformer, pulse, 350
  rotary, 574, 579
Transient operation of diodes, 91
Transient rectangular pulse, amplitude spectrum of, 103
Transient response, 65
  of capacitance, 68
  of circuit elements, 64
  of inductance, 67
  of pairs of circuit elements, 70
  of pairs in shunt, 75
  of resistance, 66
  of tube capacitance, 88
  of vacuum tubes, 80
Transit time, effect of, 90
Transmission line, 134
  constant-k artificial, 354
  input impedance of, 146
  mechanism of transmission in, 137
  methods of coupling to, 377
  parameters of, 170
  power measurements in, 164
  reflected wave at termination, 143
  uniform, 138
  wave on, 141
Transmission system, 494
Transmit-receive boxes, 395
Transmit-receive cavities, 401
Transmit-receive switches for microwave systems, 399
Transmitter, 458, 484, 491
  SCR-268, 474
Transmitting oscillators, 366
Transponder, iff, 409
Transverse electric mode, 177
  in circular guides, 189
  in rectangular guides, 183
  three-dimensional representation of, 178
Transverse magnetic modes, 190
Transverse magnetic propagation, 182
Trapezoidal wave, circuit for generating, 553
Trapping, 272, 277
  limiting angles for, 278
Trigger circuit, 331
Triode, disk-seal, 369

Triode oscillators, external-line, 367
Triple-dipole nondirectional antenna, 411
Tube, capacitance of, transient effect of, 88
in ıctive output of, 415
t-r, 400
Tuned circuit, parallel-line, 369
Tuned-cathode, tuned-grid, oscillator, 374
Tuned-grid tuned-cathode oscillator, 374
Tuners, 614
double-stub, circle diagram of, 615
slug, dielectric, 617
Tuning auxiliaries, 519
Tuning indication, i-f pickup, 519
Tuning screws, 201
Turnstiles, 410
Tuska, C. D., 7
Tuve, M. A., 7
Twists in waveguide, 204
Two-terminal-pair network, 113
Type A indication, 32, 42, 55
deflection system for, 546
modifications of, 37
Type B indicator, 33, 43
of AN/MPG-1 radar, 586
deflection system for, 546
distortion introduced by, 34
distortionless, 589
Type C indicator, 35
deflection system for, 547
distortion introduced by, 35
Type D indicator, 40
Type E indicator, 39
Type F indicator, 39
Type G indicator, 39
Type H indicator, 39
Type I indicator, 40
Type J indicator, 37, 42
deflection system for, 549
Type K indicator, 38
Type L indicator, 38
Type M indicator, 38
Type N indicator, 39
Type R indicator, 39

U

Ultrahigh-frequency oscillator circuits, 370
Unit step function, reproduction of, 121

V

Vacuum tubes, transient response of, 80
Vector potential, 221
Velocity, group, 179–181
phase, 141, 179–181
of propagation, 141
Vertical coverage diagram, 263
Video amplification, 516
linear, 338
Video amplifiers, cathode-coupled, 346
compensation of, 343
connections of, 519
equivalent circuits of, 339
low-frequency amplitude response of, 342
low-frequency performance of, 345
low-frequency phase response of, 343
phase response of uncompensated, 341
power relations in, 346
pulse transmission by, 118
series-compensated, 115
uncompensated, high-frequency amplitude response of, 340
Video coupling connection, system functions of, 106
Video detection and detectors, 516, 518
Voltage, feedback, 372
spread, 573

W

Ward-Leonard drive, 568
control amplifier for, 569
Water calorimeter, 598
Watson, C. W., 310

Watson-Watt, Sir Robert, 7, 62
Waves, paths of, refractive bending of, 271
  plane electromagnetic, 176
  sawtooth, synchronized production of, 331
  standing, 144, 165
Waveform, distortion, 94
  extended, synchronized production of, 330
  output, steady-state, computation of, 117
  spectra of, 94
Waveguides, 134
  attenuation in, 191
  attenuators, 601
  bends in, 204
  butt joint, 364
  choice of mode in, 195
  circular, attenuation in, 194
    modes in, 188
    transverse electric modes in, 189
    transverse magnetic modes in, 190
  dimensions of, 195
  excitation of, 197
  general principles, 173
  impedance matching in, 200
  mechanism of propagation in, 175
  modes in, methods of exciting, 198
  rectangular, attenuation in, 193
    geometry of, 185
    transverse electric modes in, 183
    various transverse electric modes, 187
  reflecting elements in, 200
  Smith circle diagram applied to, 200
  termination of, 197
  twists in, 204
  wavelength in, 179–181
Waveguide connectors, coaxial, 366
Waveguide corners, 204
Waveguide feeds for paraboloids, 251
Waveguide fittings, 359
Waveguide joints, 364
Waveguide radiators, 230
  gain of, 231
Waveguide rotating joint, 365
Waveguide tee, 203
Wavelength, choice of, 21
  in guide, 179–181
  in radar equation, influence of, 51
Wavemeter, cavity, 608
  coaxial, 607
  resonance, methods of indicating, 607
Weather phenomena, relation of M curves to, 278
Wert, C. A., 589
Wheeler, H. A., 344
Whinnery, J. R., 217, 290, 368
Width, beam, 52
  pulse, 299
Willemite, 539
Wilson, I. G., 626
Window, 46
Wolff, I. G., 407

X

X band, 135
$X_{m,n}$, values of, 190

Y

Young, L. C., 6

Z

Zahl, H. A., 310
Zero-error position, 569